Dynamic Bayesian smooth transition autoregressive (DBSTAR) models for non-stationary nonlinear time series

Thesis

THE OPEN UNIVERSITY

Dynamic Bayesian Smooth Transition Autoregressive (DBSTAR) models for non-stationary nonlinear time series

by

Alexandre José dos Santos, BSc, MSc

A thesis submitted in fulfilment of the requirements

for the degree of Doctor of Philosophy

in the

Department of Mathematics and Statistics

The Open University, UK

March 2014

DATE OF SUBMISSION: 15 MAY 2014

DATE OF AWARD: 29 AUGUST 2014

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

Abstract

In this thesis, Dynamic Bayesian Smooth Transition Autoregressive (DBSTAR) models are proposed for nonlinear time series, as an alternative to both the classical Smooth Transition Autoregressive (STAR) models and Computational Bayesian STAR (CBSTAR) models. DBSTAR models are autoregressive formulations of dynamic linear models based on polynomial approximations of transition functions of STAR models. Unlike classical STAR and CBSTAR models, their parameters vary in time, being suitable for modelling both global and local non-stationary processes. Since DBSTAR models are Bayesian, the models do not require extensive historical data for parametric estimation and allow expert intervention via prior distribution assessment of model parameters. Because they are analytical and sequential, DBSTAR models, respectively, avoid potential computational problems associated with CBSTAR models, such as convergence issues, and allow fast estimation of dynamic parameters sequentially in time, being thus suitable for real time applications. Proposed DBSTAR models have been applied to two data sets: the much used Canadian Lynx data set, in which the aim is to validate DBSTAR models by comparing their fitting performances with existing approaches in the literature, and a Brazilian electricity load data set, for which existing models are not suitable.

Acknowledgements

I wish to use this opportunity to express my gratitude to everyone who supported me throughout the course of this PhD project. I am thankful for their aspiring guidance, invaluably constructive criticism and friendly advice.

First and foremost I would like to thank my main supervisor Dr Alvaro Faria for encouraging my research and for allowing me to grow as a research scientist. Your guidance and insights during our weekly basis meetings were precious for this PhD project. I appreciate all your contributions of time and ideas to make my PhD experience fruitful.

I would also like to thank my co-supervisors Dr Catriona Queen and Dr Reinaldo Souza for their great ideas, suggestions, feedback and emotional support throughout the PhD. Without their co-supervision and constant help this PhD thesis would hardly have been completed.

I extend my gratitude to all members of the Statistics group at the Open University for providing an excellent working environment in various forms. Many thanks to my office mates Dr Yonas and Sofia and colleagues Dr Angela, Dr Fadl, Dr Osvaldo for being a great help in the process. The financial support from the Open University is greatly acknowledged.

I am deeply grateful to Peter whose support, encouragement and care kept me going through. My very special thanks go to Loes, Gerardo, Julia, Lilia and her family for their support, advice and encouragement. I would also like to thank my close friends Aline, Cristiane and Rosane. I greatly value their friendship and I deeply appreciate their belief in me.

Words cannot express how grateful I am to my family for all the love, encouragement, care and prayer. I have to give a special mention for the source of love, concern, support and strength all these years given by my parents Sebastiao and Janete, my sister Laura, my brother Robson and my nephew Matheus. I love you and I dedicate this thesis to my family.

Thank you everyone that I could not mention here who have encouraged and helped me in one way or another.

Contents

Abstract i

Acknowledgements ii

Table of Contents iv

List of Tables ix

List of Figures x

List of Abbreviations xiv

1 Introduction 1

1.1 Research motivation 3

1.2 Thesis outline 6

2 Literature review on STAR models 8

2.1 Important basic concepts in nonlinear time series 9

2.1.1 Nonlinearity of underlying processes 9

2.1.2 Nonlinearity in time series models 11

2.1.3 Regimes 13

2.1.4 Detecting nonlinearity in time series analysis 13

2.1.5 Non-stationarity 15

2.2 Historical remarks 16

2.2.1 Classical inference 16

2.2.2 Bayesian inference 19

2.3 Classical STAR models 22

2.3.1 Threshold Autoregressive (TAR) models 23

2.3.2 Self-Exciting TAR (SETAR) models 25

2.3.3 Smooth Transition Autoregressive (STAR) models 27

2.4 Bayesian STAR models 31

2.4.1 Bayesian TAR models 31

2.4.2 Bayesian SETAR models 34

2.4.3 Bayesian STAR models 35

2.5 Variants of STAR models 36

2.5.1 Multiple Regimes STAR models 36

2.5.2 Time-Varying STAR models 38

2.6 Summary 39

3 Dynamic Linear Models and Mathematical background 42

3.1 Dynamic Linear Models 43

3.1.1 AR models in DLM form 46

3.2 Parametric prior-to-posterior updating 47

3.2.1 Prior distributions 48

3.2.2 Filtering distributions 49

3.2.3 Smoothing distributions 53

3.2.4 Forecast Distributions 54

3.3 Bayesian model selection for determination of parameters 55

3.3.1 Bayes factor criterion 56

3.3.2 Log-predictive likelihood criterion 57

3.3.3 Log-smoothing likelihood criterion 57

3.4 Polynomial approximations 58

3.5 Spline functions 62

3.5.1 Truncated Power Basis (TPB) 63

3.5.2 B-Splines 65

3.6 Solution for overdetermined systems of polynomial equations 69

3.7 Summary 71

4 Dynamic Bayesian STAR models 73

4.1 Model definition 74

4.2 Taylor DBSTAR Models 79

4.2.1 Estimating AR and transition parameters 83

4.3 Splines DBSTAR models 89

4.3.1 Truncated Power Basis DBSTAR models 91

4.3.2 B-splines DBSTAR models 92

4.4 Discussion 94

5 Modelling further components in DBSTAR models 97

5.1 DBSTAR models for non-stationary processes 98

5.2 DBSTAR models for periodic processes 101

5.3 Incorporating predictor variables in DBSTAR models 103

5.4 Multiple Regimes DBSTAR models 106

5.5 General DBSTAR models 108

5.6 Discussion 109

6 Modelling Canadian lynx data **113**

6.1 The data set and initial data analysis 114

6.2 Formulating DBSTAR models 119

6.3 Configuring DBSTAR models 124

6.4 Comparative model analysis 129

6.5 Forecasting performance 130

6.6 Discussion 134

7 Forecasting short-term electricity load in Brazil **136**

7.1 The Brazilian electricity market 137

7.1.1 Electricity crisis in 2001-2002 138

7.1.2 The need for electricity load forecasting 140

7.2 Electricity load data set description 141

7.3 Exploratory data analysis 143

7.4 DBSTAR model formulation 151

7.5 In-sample analysis 154

7.5.1 Configuring DBSTAR models 154

7.5.2 Modelling the components of DBSTAR models 156

7.5.3 In-sample 1-step ahead forecast error diagnostic 162

7.6 Forecasting performance 164

7.7 Discussion 171

8 Conclusion **174**

8.1 Summary and main contributions 174

8.2 Future research 179

Bibliography **183**

List of Tables

6.1 Final DBSTAR model configurations for the Canadian lynx data set . 126

6.2 Model comparison - mean absolute errors (MAE) and root mean squared errors (RMSE) - DBSTAR models versus Competitors 130

6.3 Out-of-sample log-predictive likelihood (LPL) of each DBSTAR model 132

7.1 Sample division 143

7.2 Final DBSTAR model configurations for the electricity load data set . 156

7.3 Out-of-sample mean absolute percentage error (MAPE) and log-likelihood of predictive distribution (LPL) of each DBSTAR model and the random walk (RW) model 165

List of Figures

2.1 Canadian lynx time series, yearly observed from 1821 to 1934. 10

2.2 Hyperbolic tangent function as the transition function with $\gamma = 0.5$ on Panel (a) and $\gamma = 10$ on Panel (b) and, respectively, simulated time series from model (2.5) on Panels (c) and (d) 21

2.3 Logistic function $\pi(s_t, \gamma, c)$, with $\gamma = 0, 0.1, 0.5, 1, 10$ and $c = 100$, together with the indicator function $I(s_t; c), c = 100$, both as the transition function. 29

3.1 Graphical representation of the conditional independence structure of a DLM 44

3.2 Taylor series expansion of (a) degree r = 1 and (b) degree r = 3, of a logistic function around at $s_t = c$ with $\gamma = 0.3$ and $c = \bar{s_t}$ 61

3.3 (a) Piecewise polynomials of degree $r = 1$ and (b) Truncated Power Basis approximation of degree $r = 1$, of a logistic function at variable $s_t, \gamma = 0.3$ and $c = \bar{s_t}$ 65

3.4 B-spline basis functions of (a) degree $r = 1$ and (b) degree $r = 3$, of a logistic function at variable $s_t, \gamma = 0.3$ and $c = \bar{s}_t$, as well as the corresponding B-spline approximations (c) and (d), respectively. . . . 68

6.1 (a) Original Canadian lynx time series and (b) log-transformed Canadian lynx time series, yearly observed from 1821 to 1934 115

6.2 Normal Q-Q Plot of (a) original Canadian lynx time series and (b) log-transformed series 117

6.3 (a) Autocorrelation function (ACF) and (b) Partial ACF of log-transformed series 117

6.4 Periodogram of original Canadian lynx time series 118

6.5 DBSTAR model configurations structured by a classification tree . . . 121

6.6 Harmonics of (a) Model 11 and (b) Model 12 127

6.7 Observed versus fitted models - Panels (a) - (d) Taylor DBSTAR models 1, 2, 3 and 4, Panels (e) - (h) Truncated Power Basis DBSTAR models 5, 6, 7 and 8, and Panels (i) - (l) B-splines DBSTAR models 9, 10, 11 and 12 128

6.8 Observed series during out-of-sample period, along with 1-step ahead forecast means and 95% credible interval of Models 3 (a), 7 (b) and 11 (c). 133

7.1 Brazil map - Southeast and Central-West regions 138

7.2 Average hourly electricity load (MW) by days of the week plus holidays and bridge-holidays - Southeast and Central-West regions 144

7.3 Average daily electricity load in (a) June 2008 and (b) June 2009. Notice that the shape of the curve (b) changes due to holiday and bridge-holiday, as identified between the two dashed lines 145

7.4 Average weekly electricity load (MW) versus temperature in degree Celsius ($°C$) - Southeast and Central-West regions. The curve indicates a S-shape pattern rather than a U-shape for the Brazilian data, which suggests a nonlinear relationship between electricity load and temperature. 147

7.5 Average monthly electricity load in MegaWatts (MW) - Southeast and Central-West regions 148

7.6 Typical hourly electricity load curve shapes according to the season of the year, summer and winter, respectively. (a) from Sunday, 11 January 2009 to Saturday, 24 January 2009 and (b) from Sunday, 14 June 2009 to Saturday, 27 June 2009. 149

7.7 Panel (a) Month plot of average monthly electricity load in MegaWatts (MW) and panel (b) Month plot of temperature in Degree Celsius ($°C$) - Southeast and Central-West regions. The red lines are the average for the entire year. 150

7.8 Average in-sample hourly observational standard deviation estimate ($\sqrt{S_t}$) in MW by days of the week plus holidays and bridge-holidays. 157

7.9 In-sample posterior mean of coefficients associated with polynomial regression variables. 158

7.10 In-sample posterior mean of the trend component - the linear growth rate. 159

7.11 In-sample posterior mean of the periodic components to represent (a) daily cycle and (b) weekly cycle with 95% credible interval. The period in (a) is from the first hour on 11 Jan 2009 to the last hour on 17 Jan 2009, and in (b) is from the first hour on 04 Jan 2009 to the last hour on 31 Jan 2009. 160

7.12 In-sample posterior mean of the component for holidays and bridge-holidays. 161

7.13 In-sample Panel (a) histogram together with normal density function (blue curve) and Panel (b) boxplot of 1-step ahead standardised forecast residuals. 163

7.14 Out-of-sample 1-step ahead forecasting performances of Model C: Panel (a) from hour 1 on 1 June 2010 to hour 24 on 5 June 2010, Panel (b) from hour 1 on 6 June 2010 to hour 24 on 12 June 2010, Panel (c) from hour 1 on 13 June 2010 to hour 24 on 19 June 2010, Panel (d) from hour 1 on 20 June 2010 to hour 24 on 26 June 2010 and Panel (e) from hour 1 on 27 June 2010 to hour 24 on 30 June 2010. 167

List of Abbreviations

AR	Autoregressive
ACF	Autocorrelation Function
DLM	Dynamic Linear Models
CBSTAR	Computational Bayesian Smooth Transition Autoregressive
DBSTAR	Dynamic Bayesian Smooth Transition Autoregressive
LM	Lagrange Multipliers
MCMC	Markov Chain Monte Carlo
PACF	Partial Autocorrelation Function
SETAR	Self-Exciting Threshold Autoregressive
SPE	System of Polynomial Equations
STAR	Smooth Transition Autoregressive
TAR	Threshold Autoregressive

Chapter 1

Introduction

In time series analysis, values in the future depend, usually in a stochastic manner, upon the observations available in the past and the present (Box and Jenkins, 1970). Any observed time series can be viewed as a realisation of a stochastic process. This stochastic dependence may help to predict the future from the past and the present.

Let Y_t be an observational time series and X_t a latent process at time $t = 1, 2, \ldots, T$. A simple general autoregressive model of order $p \geq 1$ is defined as

$$\begin{aligned} Y_t &= X_t + \epsilon_t \\ X_t &= f(x_{t-1}, x_{t-2}, ..., x_{t-p}, \underline{\alpha}) + \eta_t \end{aligned} \tag{1.1}$$

where $f(.)$ is a function of past values of the latent process X_t, $\underline{\alpha}' = (\alpha_1, \alpha_2, \ldots, \alpha_p)$ is a p-vector with parameters associated with $x_{t-i}, i = 1, 2, ..., p$, ϵ_t is the independent and identically distributed (i.i.d.) error term of the observational time series and η_t is the i.i.d. error term of the latent process; both ϵ_t and η_t with probability distributions to be specified.

The model (1.1) is a general autoregressive model which depends upon the form of the function $f(.)$. The most popular class of linear autoregressive time series models consists of Autoregressive (AR) models by Box and Jenkins (1970), for which uncorrelated error terms across time, ϵ_t's, are normally distributed with $E(\epsilon_t) = 0$, $Var(\epsilon_t) = \sigma_\epsilon^2$ and $E(\epsilon_t \epsilon_s) = 0$, for $t \neq s$.

A linear AR model of order p, or more succinctly AR(p), is defined as a linear combination of the most p recent observed values of the time series Y_t, given by

$$Y_t = \alpha_0 + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + ... + \alpha_p y_{t-p} + \epsilon_t \quad ; \quad \epsilon_t \sim N(0, \sigma_\epsilon^2) \tag{1.2}$$

where $\alpha_i, i = 0, 1, ..., p$, is a finite set of weight parameters, ϵ_t consists of a sequence of uncorrelated random variables normally distributed with zero mean and constant variance.

The model (1.2) has no time-dependent parameters $\underline{\alpha}$ and the first and second moments are also fixed over time. That model assumes that the process remains constant around its level, represented by the current information Y_t through its immediate p past values $y_{t-1}, y_{t-2}, ..., y_{t-p}$ in a linear regression form. Such behaviour characterises the concept of stationarity (Hamilton, 1994). Allowing the parameters $\underline{\alpha}$ to change over time, turns the model (1.2) to a non-stationary model.

Although linear stationary AR models have been widely applied to time series in different areas, limitations in their parametric inference have been pointed out on both classical and Bayesian methodologies and extensions have been proposed in the literature. A notably extensive class of time series models can be derived from model (1.1) by replacing the right-hand side by an arbitrary estimable function

$f(.)$ to obtain some nonlinear autoregressive models, as described in the following Chapter 2.

1.1 Research motivation

Beyond the linear domain, there are many nonlinear forms to be explored in model (1.1) in order to deal with some complications, which include asymmetric cycles (slow increases and fast decreases, or vice-versa, within a period of time, of a time series), nonlinear relationship between lagged variables, among others.

Recently, an explosion of statistical models to analyse and forecast time series presenting nonlinear complexities have been proposed in the literature. Classical Smooth Transition Autoregressive (STAR) models of Chan and Tong (1986a) and Computational Bayesian STAR (CBSTAR) models of Lopes and Salazar (2005) belong to the class of nonlinear autoregressive models which have been extensively applied to time series in different areas. These powerful families of models are suitable, practical and quite flexible for representing the underlying nonlinear processes. The main areas of applications are macroeconomics, finance and energy - just to name a few. In those areas, it is crucial to investigate data presenting nonstandard features.

Certain aspects of the processes, such as asymmetric cycles, can be described by STAR-type models. However, some critical limitations were identified by investigating both the classical STAR and CBSTAR models as well as their variants. Generally, they require modification of the data set, such as using a Box-Cox transformation, during exploratory analysis in order to remove some problematic comportment, such as non-stationarity. In addition, both the classical STAR and

CBSTAR models, as well as their variants, require large data sets to estimate the parameters. However, there are some applications in which the data are collected in wide time intervals, such as yearly data sets, say, to have enough data in hands to apply any of those models. In these cases, it takes over decades to collect data in order to have a large number of observations so these models can be applied.

On the other hand, data sets collected in very short periods of time may cause computational problems for existing models due to the large amounts of data, known as real-time or high frequency data sets. For instance, for data observed minute-by-minute, every half-hour, or hourly, decisions have to be made quickly and the model should estimate the parameters in real-time, practically.

Dynamic Bayesian Smooth Transition Autoregressive (DBSTAR) models are proposed in this thesis for non-stationary nonlinear time series as alternatives to both the classical STAR and CBSTAR models. DBSTAR models are appropriate for modelling nonlinear processes and can accommodate some unconventional and remarkable patterns, such as asymmetric cycles. In fact, the classical STAR and CBSTAR models can accommodate some of those patterns. However, DBSTAR models can be applied without the need for transforming the data, as we shall see in the application chapters, in cases of nonlinearity together with any (or all) of the following: non-stationarity, seasonality, asymmetric cycles and heteroscedasticity.

Tsay (1991) mentioned "the collection of nonlinear time series models is so vast that it is too much to expect that a single class of models is capable of capturing most of the observed nonlinear phenomena". However, the non-standard features mentioned above can be accommodated in our approach just by adding compo-

nents into the model formulation. Accordingly, DBSTAR models bring simplicity, feasibility and interpretability to the underlying processes.

Since DBSTAR models are Bayesian, either small or large historical data sets can be used for parametric estimation. Appropriate prior distributions should be chosen for the parameters. In addition, the models allow expert intervention via prior distribution assessment of model parameters. Moreover, DBSTAR models are estimated sequentially in time, which allows fast estimation of the dynamic parameters, being thus suitable for real-time applications.

Unlike the classical STAR and CBSTAR models, DBSTAR models allow both the observational and the parameter variances to vary in time, being thus suitable for modelling heteroscedastic processes. DBSTAR models incorporate a variance discounting technique, assuming that the variability may change over time but only slowly and steadily.

Since DBSTAR models may adapt to the data sequentially over time accommodating changes in the parameters at each time, the assumption of both local and global stationarity in the proposals can be straightforwardly relaxed, just by adding appropriate components related to them into the models' structure.

Once a DBSTAR model has been properly formulated, it can be used for various purposes, such as understanding and interpreting the mechanisms that generated the data and forecasting future events. We applied different formulations of DBSTAR models to the well-known Canadian lynx data set. The aim is twofold. On the one hand, we aim to validate the proposed DBSTAR models by comparing their performances against the performances of both the classical STAR and CBSTAR

models. On the other hand, the aim to illustrate the extra features we can achieve by adopting sequential models with dynamic parameters.

To illustrate a case in which the existing classical STAR and CBSTAR models are not appropriate, i.e., a non-stationary nonlinear process requiring fast sequential computations, an advanced formulation of the DBSTAR model is proposed and applied to a series of hourly Brazilian electricity load.

1.2 Thesis outline

This thesis presents in Chapter 2 a description of developments of existing nonlinear autoregressive models, in particular the classical and Bayesian STAR models and their variants.

Chapter 3 describes Dynamic Linear Models (DLMs), which are a key approach for the proposed models in this thesis, together with the parametric prior-to-posterior updating methods, a Bayesian model selection criterion and the mathematical background used for the development of DBSTAR models reported in subsequent chapters. Polynomial approximations methods, with a focus on Taylor series approximation, and splines basis functions, both methods for approximating transition functions at the model formulation stage, are described. A review of the Lagrange multipliers method for solving overdetermined systems of polynomial equations is also reported.

Chapter 4 defines and describes the proposed DBSTAR models based on both Taylor series approximations and splines basis functions for approximating the transition functions. The methodological advantages relative to existing methods (de-

scribed in Chapter 2) will be emphasized. That chapter will also include the preservation of parameter interpretability via a solution of overdetermined systems of polynomial equations and discussions.

Chapter 5 describes DBSTAR model extensions, such as modelling of processes with trend, cycle, seasonality, as well as the inclusion of predictor variables and multiple regimes modelling within the DBSTAR framework.

Chapter 6 reports the application of proposed DBSTAR models to the Canadian Lynx data set. One of the aims is to validate DBSTAR models by comparing their fitting performances with those of existing methods in the literature.

Chapter 7 reports a second application of proposed DBSTAR models, which refers to an hourly Brazilian electricity load data set. This is an application that will show a case in which existing approaches are not suitable for adoption.

Conclusion and future research are presented in Chapter 8. That chapter summarises the main overall results of the proposed models as well as directions for further research which include extensions to the model for more than 2 regimes, multivariate DBSTAR formulations and the use of particle filtering for sequential parametric estimation, based on simulation methods for more precise posterior and predictive distributions.

Chapter 2

Literature review on STAR models

Autoregressive models which are key ingredients for the DBSTAR proposed in this thesis are described in detail in this chapter. It summarises the main developments in nonlinear autoregressive models. The focus is on the principal results on both the classical and Bayesian STAR models that more closely relate to DBSTAR models.

It is worth noting that this thesis concentrates on stochastic situations only, not covering for instance chaotic series or artificial intelligence approaches such as neural networks. Only discrete univariate time series models will be considered for which discrete time-intervals were specified during data collection such as hours, days, months, years, and so forth.

In the following sections, the notation adopted throughout this thesis will be introduced. A review on classical and Bayesian TAR and STAR models are covered in Sections 2.3 and 2.4, respectively. The fitting performance of some of these models are compared with the fitting performance of DBSTAR models in Chapter 6.

2.1 Important basic concepts in nonlinear time series

In this section, important basic concepts in nonlinear time series are described, as they will be used hereinafter.

2.1.1 Nonlinearity of underlying processes

Many studies have detected nonlinearity of underlying processes by investigating some non-standard behaviours of the time series under analysis. An illustration of such behaviours is given in this section, with a focus on asymmetric cycles.

In time series analysis, cycles have repetitive forms measured within any period of time (Pole *et al.*, 1994). For instance, a cycle can be identified hourly, daily, weekly, and so forth. A specific type of cycle is called seasonality, when the periodic repetition occurs within one-calendar year, such as monthly (12 periods), quarterly (4 periods), etc.

A distinctive non-standard behaviour that can be found in time series data sets is the asymmetric cycle, i.e., a time series that shows a slow increase and a fast decrease, or vice-versa, within a period of time. The rise/decline period of the cycle exceeds the descent/ascent period, characterising an asymmetric periodic behaviour in the frequency. This asymmetry is referred to as amplitude-frequency dependency (Tong and Lim, 1980).

Such cyclic oscillations with changes in the frequency and/or in the amplitude are investigated in different areas of research. For instance, Tong and Lim (1980) in-

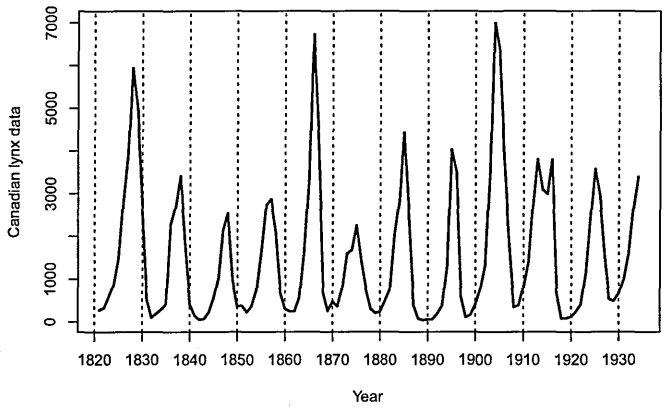


Figure 2.1: Canadian lynx time series, yearly observed from 1821 to 1934.

vestigated nonlinearity, related to amplitude-frequency dependency, in different data sets. Using the Canadian lynx data set illustrated in Figure 2.1, they concluded that the process presents cycles of approximately 10 years with varying frequency and amplitude. The rise period of the cycle exceeds the descent period, characterising an asymmetric periodic behaviour in the frequency. Notice the dashed vertical lines in the graph. Generally, it takes about six years to rise and just four years to descend. Similar analysis was carried out using a time series of sunspot. They also concluded that these were asymmetric cycles present in this process. The rise period of the cycle is shorter than the descent period, which runs in an opposite direction to that of the Canadian lynx.

Therefore, linear models would not be suitable for modelling such processes. The understanding of processes with asymmetric cycles would be more appropriate with the use of a much wider class of nonlinear models. Priestley (1988) stated: "if the series under study exhibits a cyclical phenomenon, but with asymmetrical cycles, then a nonlinear model may well provide a more satisfactory description of

the data."

Veselovsky and Tarsina (2002) also analysed nonlinearity of the solar cycles using a monthly averaged sunspot numbers data set. They state that "stronger solar cycles are shorter, as a rule". Their results showed "non-sinusoidal time dependence with shorter rising phases and longer declining phases of moderate and strong cycles".

Zhang and Duda (2013) examined nonlinearity of internal tidal waves provoked at the shelf break. Nonlinear effects in this application cause additional superharmonic and subharmonic internal waves, with the presence of a stronger nonlinearity on subharmonic waves. Subharmonics are oscillations at a fraction of the frequency and superharmonics are oscillation at a multiple of the frequency (Tong and Lim, 1980). Both types of harmonics also describe asymmetric cycles.

A nonlinear time series model should then reflect these underlying characteristics present in such processes.

2.1.2 Nonlinearity in time series models

The AR model in equation (1.2) present linearity in their parameters. Linearity, in this case, refers directly to the physical structure of the model in which the quantities to be estimated, the parameters α_i, appear linearly in the formulation (Kutner *et al.*, 2005). Other physical model structures of the general model in equation (1.1), such as polynomials with power functions or other forms, also present linearity in their parameters, for instance,

$$Y_t = \alpha_0 + \alpha_1 y_{t-1} + \alpha_2 y_{t-1}^2 + ... + \alpha_r y_{t-1}^r + \epsilon_t \tag{2.1}$$

where $\alpha_i, i = 0, \ldots, r$, are the coefficients of the polynomials. Notice that for $r = 1$, the model (2.1) becomes an AR(1), however, $r = 2$ describes a parabolic model and $r = 3$ represents an S-shape curve model. Nonetheless, those models also fall into a linear category of the above meaning. The possible reason for that could be an investigation of their first derivatives. Notice that the first derivative of the model (2.1) with respect to any of the parameters α_i is independent of that parameter. On the other hand, if a parameter, say γ, is introduced to the model, as follows,

$$Y_t = \alpha_0 + \gamma\alpha_1 y_{t-1} + \gamma\alpha_2 y_{t-1}^2 + \ldots + \gamma\alpha_r y_{t-1}^r + \epsilon_t \tag{2.2}$$

the first derivative of the new model (2.2) with respect to any of the parameters α_i is not independent of γ, or the first derivative with respect to γ is not independent of any α_i either. This would characterise nonlinearity in the parameters.

Using geometric concepts, Bates and Watts (1980) and Hamilton *et al.* (1982) stated that, intrinsic nonlinearity is inherent to the models (1.1) with nonlinear parameter effects as above mentioned in model (2.2). A suitable linearisation of nonlinear models with particular choice of the parameters, such as the Taylor series approximation, produces satisfying reparametrisation of the models, as their mathematical forms change from the form of model (2.2) to the form of model (2.1), say, which remove the parameter-effects nonlinearity. So a nonlinear model can be linearised by a Taylor series approximation.

2.1.3 Regimes

In nonlinear autoregressive time series analysis, the relationship between Y_t and its past Y_{t-p} is analysed by attributing a different linear autoregression to each two distinct regions, called regimes. Basically, data are divided into sub-sets represented by linear autoregressive models.

The central idea of autoregressive models using regimes in their structure is to approximate a nonlinear relationship between Y_t and its past by a piecewise linear autoregression model (1.2) in each one of the regimes, i.e., there is a change in the linear coefficients according to an assumed value of an external variable s_t, denoted as the transition variable. So this variable is responsible for representing the information that is causing nonlinearity to the process. Therefore, autoregressive models with regimes are partly linear since they are linear in the parameters.

Either a complete switch or a smooth transition between linear autoregressive models may occur. This characterises one type of nonlinear autoregressive models (Tong, 1978), as the combination of these individual autoregressive models is seen as a piecewise linear model which assumes the role of imitating the nonlinearity present in the process.

2.1.4 Detecting nonlinearity in time series analysis

In general, time series components such as trend, cycle and seasonality, as well as heteroscedasticity can be straightforwardly diagnosed by graphical analysis. However, certain characteristics such as asymmetric cycles and nonlinear relationships between variables are less straightforward to pick out.

In time series analysis, it can be difficult to identify certain patterns that indicate nonlinearity by visual inspection alone. In those cases, more detailed investigation is usually required to detect nonlinearities in the underlying process. Hypothesis testing has been used to show whether linear or nonlinear models best represent the underlying process. There have been many approaches proposed in the literature. Usually, the null hypothesis is set to linearity against the alternative hypothesis of nonlinearity.

Tsay (1989) proposed hypothesis tests for testing linearity against threshold autoregressive models whereas Luukkonen *et al.* (1988) and Teräsvirta (1994) proposed tests for linearity against smooth transition autoregressive models. Both approaches have been widely used and extended since. More general tests can be also found in the literature, e.g., Tsay (1986), Chan and Tong (1986b) and Tsay (1991) proposed quite general procedures to detect various nonlinearities in univariate time series.

However, misidentification problems may occur when using these approaches. For identifying nonlinearity in time series, the existing methods first specify some linear and nonlinear models as for the null and alternative hypothesis, respectively, and then conclude for evidence of linearity or nonlinearity for the process. Notice that, in fact, the existing approaches only test linear AR models against nonlinear AR models, in an attempt to identify the characteristics of linearity or nonlinearity. For example, the nonlinear model under investigation, such as a classical STAR model defined in Section 2.3, is set as an alternative hypothesis together with the null hypothesis of a linear model, e.g. a linear AR model. A problem associated with this hypothesis testing could be that the evidence for nonlinear structure in the

data may be coming from some aspects of the data not following the specific linear AR model set as the null hypothesis, but from another linear model. In such case, the conclusion for modelling the data using a STAR model would be inaccurate.

In this thesis nonlinearity is not identified via hypothesis tests. Instead, an exhausting exploratory analysis is carried out based on either the concept of asymmetric cycles, i.e., trying to identify them, or the study of the relationship between the dependent time series and an external one.

2.1.5 Non-stationarity

There are different concepts of non-stationarity in the literature. The two most used are the weakly non-stationarity and the strictly non-stationarity (Hamilton, 1994). Moreover, two subgroups of those can be identified, locally non-stationarity and globally non-stationarity (Priestley, 1988).

The process for Y_t is weakly non-stationary if its moments depend on time t, for all $t = 1, \ldots, T$. Generally, the first two moments are investigated. On the other hand, the process is strictly non-stationary if, for any values of $j_1, \ldots, j_n$, the joint distribution of $Y_t, Y_{t+j_1}, \ldots, Y_{t+j_n}$, depends on both the delays $(j_1, \ldots, j_n)$ separating the dates and on the date t itself, for all $t = 1, \ldots, T$.

A locally non-stationary process is identified by splitting the time axis, representing the parameter space, into a large number of small divisions and checking whether either weakly or strictly non-stationarity is detected in any division. Besides, a globally non-stationary process is identified using the whole period $t = 1, \ldots, T$, i.e., without splitting the parameter space to check the presence of non-stationarity.

Further, if the vector of parameters $\underline{\alpha}$ in (1.1) changes over time, i.e., time-dependent parameters, the model becomes a non-stationary model. Notice that, in this thesis, the terminology "non-stationarity" itself means locally weakly non-stationarity, unless stated otherwise.

An example of a model that represents a stationary process is the model (1.2), which has no time-dependent first and second moments as well as fixed parameters over time. That model assumes that the process remains constant around its level, represented by the current information Y_t through its immediate p past values $y_{t-1}, y_{t-2}, ..., y_{t-p}$ in a linear regression form. If the parameters $\underline{\alpha}$ are allowed to change over time, i.e., $\underline{\alpha}_t, t = 1, \ldots, T$, the model becomes a non-stationary AR model, hence suitable for modelling processes with trend.

2.2 Historical remarks

Although linear stationary AR models are computationally simple and have been widely applied to time series in different areas, it soon became clear that many time series problems required a more sophisticated model. Limitations in the AR parametric inference have been pointed out in both classical and Bayesian methodologies, as described in this section and extensions have been proposed in the literature, as reported in Sections 2.3, 2.4 and 2.5.

2.2.1 Classical inference

On the classical inference side, the first studies that detected limitations of linear models were conducted in The Fifties. Moran (1953) statistically analysed the well-

known Canadian lynx data set, illustrated in Figure 2.1, using linear AR models in an attempt to represent the process of population dynamics. In that paper, attention was drawn to the residuals that showed a curious behaviour. Two distinct groups of residuals were formed, one that the linear model produced from the sample points greater than the mean and another from the sample points smaller than the mean. It was detected that the former were significantly small compared with the latter. It was then concluded that, since there are some patterns leftover in the residuals which were not captured by the model, the underlying process is asymmetric and should not be modelled by a linear AR model.

Whittle (1954) analysed a water level time series collected in a rock channel on the New Zealand coast. This small channel contains atypical depths at shallow water and opens into partly protected water of a bay. In addition, it runs parallel to a corridor of breakers outside but it is disconnected from them by a strip of rocks. A significant arithmetical relationship was detected among the periods of the pronounced peaks of the waves. This relationship produced some evidence for a rather convincing nonlinear structure, which is beyond the domain of linear models. Hence, a more sophisticated model than the linear AR model is required.

Research in detecting changes in the parameters was discussed by Page (1954) when all observations were divided into sub-samples, each of which refers to a random sample from the same distribution but regarding different parameter values. Hypothesis tests were proposed by Page (1955, 1957) to detect that, given T independent observations $x_1, \ldots, x_T$ observations, they come from the same population with distribution function $F(x|\theta)$ against the alternative that the first τ observations

come from a distribution function $F(x|\theta_1)$ and the remaining $T-\tau$ observations come from a distribution function $F(x|\theta_2)$, with unknown τ, $0 < \tau < T$ and $\theta_1 \neq \theta_2$. The main concern in splitting the sample into two is to detect changes in the parameter values from θ_1 to θ_2.

Based on those findings, Quandt (1958) suggested a method of estimating a linear regression system obeying two regions, referred to as regimes, i.e., the method attributes a different linear regression to each two distinct regimes. Since the analysis is not affected by the number of explanatory variables, X_t, in the regression, for a time t, with $t = 1, 2, ..., T$, define the case of two relationships between a dependent variable, Y_t, and one explanatory variable X_t:

$$Y_t = \phi_{0,1} + \phi_{1,1}X_t + \epsilon_{1,t}, \quad t = 1, \ldots, \tau \tag{2.3}$$

$$Y_t = \phi_{0,2} + \phi_{1,2}X_t + \epsilon_{2,t}, \quad t = \tau + 1, \ldots, T \tag{2.4}$$

where for $i = 1, 2$, $\phi_{0,i}$ and $\phi_{1,i}$ are the model coefficients and $\epsilon_{i,t}$ is independent normally distributed error terms with zero mean and constant variances σ_i^2. In addition, assume $\epsilon_{i,t}$ is also independent of the explanatory variable X_t. For these two relationships, a total of T observations is considered and divided into two regimes: equation (2.3) for the first $t = 1, \ldots, \tau$ observations and equation (2.4) for the remaining $t = \tau + 1, \ldots, T$. The identified problem was to estimate the time τ at which the model switches from one regime to the other. This approach considered one likelihood function for the first τ observations and another for the remaining $T - \tau$ observations. This method calculates the value of the likelihood function for

all possible values of τ, considering the entire sample. The value of τ is then selected as the one that maximises the likelihood estimate. Then, given this highest value of τ, the maximum likelihood estimates for ϕ_{01}, ϕ_{02}, ϕ_{11}, ϕ_{12} and σ_i^2 are obtained in the usual way.

In The Sixties, extensions of the methods previously mentioned were proposed. Quandt (1960) published some advances on his research by testing the hypothesis that no switch occurred in the parameters of a linear regression system against the alternative that one switch took place. Furthermore, Robison (1964) and Hinkley (1969) proposed maximum likelihood methods for the intersection of two regressions whereas Hudson (1966) proposed least squares methods. Both approaches estimate the parameters when the time τ is unknown.

More research in classical AR models after The Sixties are described in Section 2.3.

2.2.2 Bayesian inference

Still in The Sixties, Bayesian methods in this area started to appear in the literature. Some studies reviewed in this section pointed out the difficulty of computing the time τ, at which the model switches from one regime to the other. Those concluded that neither maximum likelihood nor least squares methods are efficient estimators for τ. Therefore, probability distributions were given to τ and approximation inference methods were developed for switching regression problems.

From a Bayesian point of view, Chernoff and Zacks (1964), Kander and Zacks (1966) and Mustafi (1968) proposed methods to extend the works of Page (1954,

1955, 1957) from the classical side. The Bayesian approaches explored this area by (i) giving an appropriate prior probability distribution to the time points of change τ, (ii) investigating the amount of change in the parameters since appropriate prior probability distributions were also given to them, (iii) proposing Bayes estimators to the parameters, and (iv) designing simpler, more appropriate and more powerful alternative techniques to the hypothesis tests.

At the same time, Box and Tiao (1965) investigated non-stationary time series with a possible shift in the level after the occurrence of an event at a certain time τ and proposed a Bayesian approach to make inference about it.

In the Seventies, Bacon and Watts (1971) proposed a Bayesian estimation approach for detecting the switch τ between two straight lines. Their proposal extends the previous work in this area because it can accommodate either an abrupt switch or a smooth transition from one regime to the other, defined as

$$Y_t = \phi_0 + \phi_1(X_t - c) + \phi_2(X_t - c)\pi\{(s_t - c)/\gamma\} + \epsilon_t \tag{2.5}$$

where $\pi\{.\}$ (short for transition) is a function which produces the transition from one regime to the other, with γ as a transition parameter, s_t as a transition variable, i.e., an external variable that specifies the transition between regimes, c as a threshold point, i.e., the point when a transition between regimes occurs and $\epsilon_t \sim N(0, \sigma^2)$. This method can be used whether or not the nature of the transition (either smooth or abrupt) is known *a priori*. Furthermore, under certain conditions, many transition functions $\pi\{.\}$ could be adopted, such as the cumulative distribution function of a symmetric probability density function or the hyperbolic tangent function, as

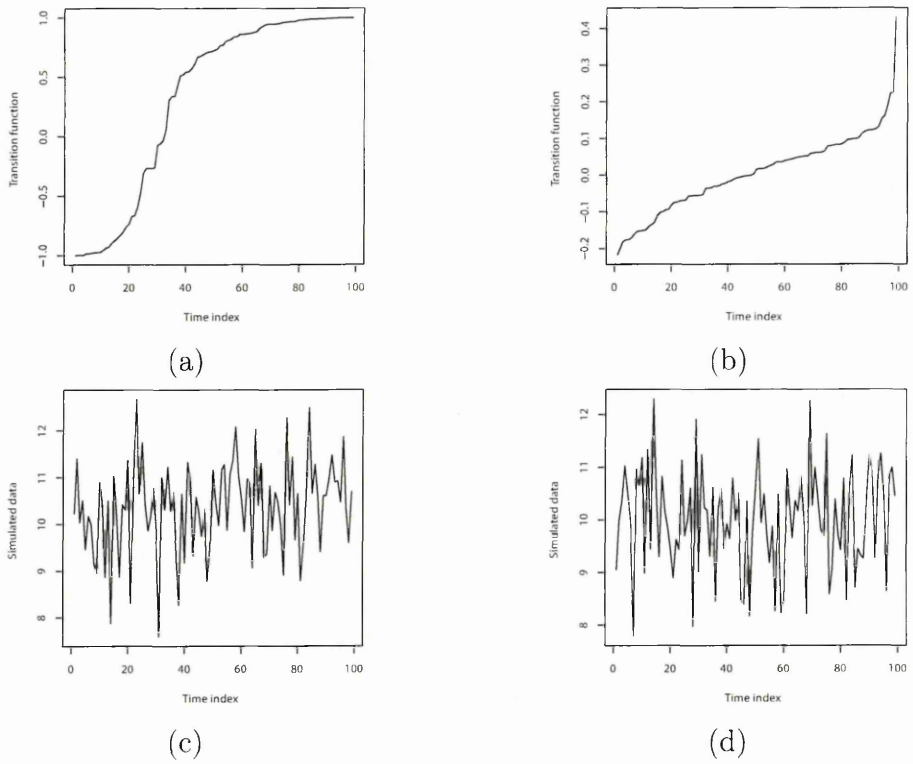


Figure 2.2: Hyperbolic tangent function as the transition function with $\gamma = 0.5$ on Panel (a) and $\gamma = 10$ on Panel (b) and, respectively, simulated time series from model (2.5) on Panels (c) and (d)

variability in the data mask any difference introduced by the choice of them.

Consider the hyperbolic tangent function expressed as

$$tanh(x) = \frac{sinh(x)}{cosh(x)} = \frac{1 - e^{-2x}}{1 + e^{-2x}} \tag{2.6}$$

where $tanh(x)$ is the hyperbolic tangent of x, $sinh(x)$ is the hyperbolic sine of x and $cosh(x)$ is the hyperbolic cosine of x. Figure 2.2 illustrates the hyperbolic tangent function of $(s_t - c)/\gamma$ used as the transition function between regimes for $\gamma = 0.5$ on Panel (a) and $\gamma = 10$ on Panel (b), with threshold $c = 50$, together with simulated time series from model (2.5) on Panels (c) and (d), respectively, with $\phi_0 = 0.1, \phi_1 = -0.2, \phi_2 = 0.3$ and $\epsilon_t \sim N(0, 1)$. Notice that when $\gamma = 0.5$, the transition between

regimes is smooth, however when γ increases to 10 the hyperbolic tangent function becomes smoother and, consequently, the transition between regimes occurs slower. This behavior is reflected in the simulated time series; the larger the γ the more oscillations the time series have. Notice also from Panels (c) and (d) that the model (2.5) should be used to represent homoscedastic stationary nonlinear processes only.

Ferreira (1975) proposed Bayesian methods extending the work of Quandt (1958) for analysing the time τ that a switch from one regime to another occurs, for a known number of regimes. Prior probability distributions were given to τ with Bayesian estimation methods to obtain posterior probability distributions for the parameters. This was slightly more complicated than the maximum likelihood method but produced very similar results when applied to the same data set as in Quandt (1958).

As it can be noticed from the above in this section, the historical remarks on nonlinear autoregressive models show that the early studies conducted in this area are based on limitations of linear AR models. All the above surveyed works from both the classical and Bayesian inference sides motivated the subsequent works, described in the following Sections 2.3, 2.4 and 2.5.

2.3 Classical STAR models

In spite of the vast applicability and dominance of linear models, it had long been noticeable that there are many aspects of time series in different fields which cannot properly be analysed within the traditional framework of linear autoregressive time series models. However, extensions of those models have been proposed in the literature. In this section, some classical nonlinear autoregressive models are introduced

and their properties are discussed. Some identified issues are also presented.

2.3.1 Threshold Autoregressive (TAR) models

Some limitations in linear AR models for stationary processes by Box and Jenkins (1970) motivated the development of Threshold Autoregressive (TAR) models, with a focus on applications where there are different frequencies of oscillations for different amplitudes: this is fundamentally a nonlinear phenomenon. This is just one of the observed phenomena which indicated the need to move beyond linear models.

The first ideas about TAR models were published by Tong (1977) in a contribution to the discussion of the paper by Lawrance and Kottegoda (1977) and then defined in more details in Tong (1978, 1983, 1990) as alternative models for describing non-standard periodic time series that cannot be captured by linear time series models. Analysing the residuals gives evidence that TAR models provide a pertinent description of the data.

Let Y_t be a univariate time series, for $t = 1, 2, \ldots, T$. TAR models with autoregressive order p are defined as:

$$Y_t = \begin{cases} \alpha_{0,1} + \sum_{j=1}^{p} \alpha_{j,1} y_{t-j} + \epsilon_{1,t}, & \text{if } s_t \leq c \\ \alpha_{0,2} + \sum_{j=1}^{p} \alpha_{j,2} y_{t-j} + \epsilon_{2,t}, & \text{otherwise} \end{cases} \tag{2.7}$$

where $\alpha_{0,i}, \alpha_{1,i}, ..., \alpha_{p,i}$, i = 1, 2, are the coefficients associated with the autoregressive part of the model, $\epsilon_{i,t} \sim N(0, \sigma^2)$ are the error terms, s_t is a transition variable, i.e., an external variable that specifies the transition between two regimes and c is a parameter called threshold that indicates when the transition occurs.

TAR models can be seen as piecewise linear time series models and they emanate as a natural candidate of $f(.)$ in model (1.1), i.e., a nonlinear function of past values of Y_t. The main idea is to represent the nonlinearity present in the process as locally stationary linear autoregressions by attributing a different linear autoregressive model to each of the two distinct regimes, based on the assumed value of a candidate transition variable. Their success partially lies in their simplicity in terms of both model-fitting and, perhaps more importantly, model-interpretation.

With Gaussian distributions given to the error terms, the likelihood function can be easily written down and the maximum likelihood estimates of the parameters derived. The parameters can also be estimated by ordinary least square methods. Therefore, TAR models work in a similar way as linear AR models, since they are piecewise linear models. The AR order p and the threshold parameter c can be chosen by using Akaike's Information Criteria - AIC (Akaike, 1974).

The model in (2.7) can be rewritten in a vectorial form, such as

$$Y_t = \underline{z}_t'\underline{\alpha}_1 I(s_t; c) + \underline{z}_t'\underline{\alpha}_2[1 - I(s_t; c)] + \epsilon_t \tag{2.8}$$

where $\underline{\alpha}_i, i = 1, 2$, is a $(p+1)$-vector with autoregressive coefficients, $\underline{z}_t' = (1, y_{t-1}, \ldots, y_{t-p})$ contains the lagged values of the variable Y_t, and $I(.)$ is an indicator (or step) function defined by:

$$I(s_t; c) = \begin{cases} 1, & \text{if } s_t \leq c \\ 0, & \text{otherwise} \end{cases} \tag{2.9}$$

The use of an indicator function as the transition function allows TAR models to abruptly switch from one AR model (regime 1) to another (regime 2). That

function specifies an abrupt switch at the time the transition variable assumes the value $s_t = c$.

A problem with the models' structure is that the likelihood functions present discontinuities that affect statistical inference due to the abrupt switching between regimes (Teräsvirta, 1994). They also require large data sets to estimate the parameters due to the use of maximum likelihood or ordinary least square estimation methods. Thus, they are not useful for real-time application. In addition, these models require the parameters, including the variance, to be fixed over time, and so are not suitable for heteroscedastic processes. Consequently, TAR models require locally stationary data sets (locally in each regime) for parameter estimation, and hence are not suitable for processes presenting a long-term stochastic (upward or downward) trend.

2.3.2 Self-Exciting TAR (SETAR) models

A question that rises when using TAR models is *how to choose the transition variable* s_t*?*. Tong and Lim (1980) defined then Self-Exciting Threshold Autoregressive (SETAR) models. These models determine the transition between two regimes by taking the transition variable s_t in equation (2.9) as a lagged value of Y_t itself, hence the self-exciting part of the name. The ability of self-excitements of the TAR family were shown in the role of amplitude-frequency dependency (Tong and Lim, 1980), for example.

SETAR models are defined as,

$$Y_t = \underline{z}_t'\underline{\alpha}_1 I(y_{t-d}; c, d) + \underline{z}_t'\underline{\alpha}_2[1 - I(y_{t-d}; c, d)] + \epsilon_t \tag{2.10}$$

where d is a delay parameter introduced to the model which determines the lag y_{t-d} of the transition variable that is provoking an abrupt change in the process. Notice that the indicator function (2.9) is adapted to $I(y_{t-d}; c, d)$ as

$$I(y_{t-d}; c, d) = \begin{cases} 1, & \text{if } y_{t-d} \leq c \\ 0, & \text{otherwise} \end{cases} \tag{2.11}$$

Like TAR models, either maximum likelihood or ordinary least squares methods can be used for estimating the parameters and the AIC can also be used to select the AR order p, the threshold parameter c and the delay parameter d.

To illustrate this idea, we report a SETAR model applied to the Canadian lynx data set by Tong and Lim (1980):

$$\hat{Y}_t = \begin{cases} 0.5239 + 1.0359y_{t-1} - 0.1756y_{t-2} + 0.1753y_{t-3} - 0.4339y_{t-4} + \\ \quad +0.3457y_{t-5} - 0.3032y_{t-6} + 0.2165y_{t-7} + 0.0043y_{t-8}, & y_{t-2} \leq 3.1163 \\ 2.6559 + 1.4246y_{t-1} - 1.1618y_{t-2} - 0.1094y_{t-3}, & y_{t-2} > 3.1163 \end{cases}$$

This estimated SETAR model can be interpreted as follows (recall that the concept of amplitude-frequency dependency has been described in Section 2.1.1): the lower regime, that is, $y_{t-2} \leq 3.1163$, corresponds to the ascent period of the cycle that, generally, takes a long time to reach highest amplitudes. The linear AR model

attributed to this regime is $\hat{Y}_t = 0.5239 + 1.0359y_{t-1} - 0.1756y_{t-2} + 0.1753y_{t-3} - 0.4339y_{t-4} + 0.3457y_{t-5} - 0.3032y_{t-6} + 0.2165y_{t-7} + 0.0043y_{t-8}$. Notice that the slow rise period of the cycle depends on observations of the previous 8 years (autoregressive order $p = 8$). On the other hand, the upper regime, that is, $y_{t-2} > 3.1163$, corresponds to the descent period of the cycle that, usually, takes a short time to reach the lowest amplitudes. The linear AR model attributed to this regime is $\hat{Y}_t = 2.6559 + 1.4246y_{t-1} - 1.1618y_{t-2} - 0.1094y_{t-3}$. In this case, since there is a fast decrease in the cycle, observations from the past 3 years only (autoregressive order $p = 3$) are needed. In ecology, they name the lower regime as population increase and the upper regime as population decrease phases, which can only be reflected by a nonlinear model. For further interpretations, see Stenseth *et al.* (1998).

SETAR models have also been well-used for many other time series data. They are particular cases of TAR models when the indicator transition variable assumes a past value of the dependent endogenous time series. Since the use of an indicator function still applies, similar discontinuity problems that affect statistical inference due to the abrupt switching between regimes in TAR models are also detected for SETAR models. In addition, the other limitations of TAR models are also observed in SETAR models.

2.3.3 Smooth Transition Autoregressive (STAR) models

A generalisation of TAR and SETAR models incorporating a smooth monotonically increasing transition between the two regimes was proposed by Chan and Tong (1986a) called Smooth Transition Autoregressive (STAR) models. One of the great

advantages in using models with smooth transition is the possibility of specifying the transition from one regime to the other in order to avoid the problem of having an abrupt threshold between them, i.e., the discontinuities in the conditional mean which are implicit in TAR and SETAR models can be avoided.

The mathematical expression representing a STAR model of autoregressive order p with two regimes is given by

$$Y_t = \underline{z}_t'\underline{\alpha}_1\pi(s_t;\gamma,c) + \underline{z}_t'\underline{\alpha}_2[1-\pi(s_t;\gamma,c)] + \epsilon_t \quad ; \quad \epsilon_t \sim N(0,\sigma^2) \tag{2.12}$$

where for $i = 1, 2$, $\underline{\alpha}_i' = (\alpha_{0i}, \alpha_{1i}, \ldots, \alpha_{pi})$ are $(p+1)$-dimensional vectors with elements α_{ji} $(j = 0, 1, \ldots, p)$ representing autoregressive coefficients associated with each component j of the AR regime i; $\underline{z}_t' = (1, y_{t-1}, \ldots, y_{t-p})$ is a $(p+1)$-dimensional vector; $\pi(.)$ is a nonlinear smooth transition function in the range $[0, 1]$ with parameters $\gamma \in \mathbb{R}^+$ and $c \in \mathbb{R}$ defined as smoothing and threshold, respectively, and $s_t \in \mathbb{R}$ defined as a transition variable (usually in practice being either an external variable or lagged dependent variable y_{t-d}, where d is a delay parameter, as in the previous Section 2.3.2), and ϵ_t is independent and identically normally distributed.

In the literature, we can find STAR models with different transition functions such as the first- and second-order logistic functions, the exponential function (Teräsvirta, 1994), and others such as the cumulative distribution function (Chan and Tong, 1986a) and the hyperbolic tangent function (Bacon and Watts, 1971). See Teräsvirta (1994) for more details on alternative transition functions. The most commonly used transition function, and the one we will be adopting in Chapters 6 and 7, is the first-

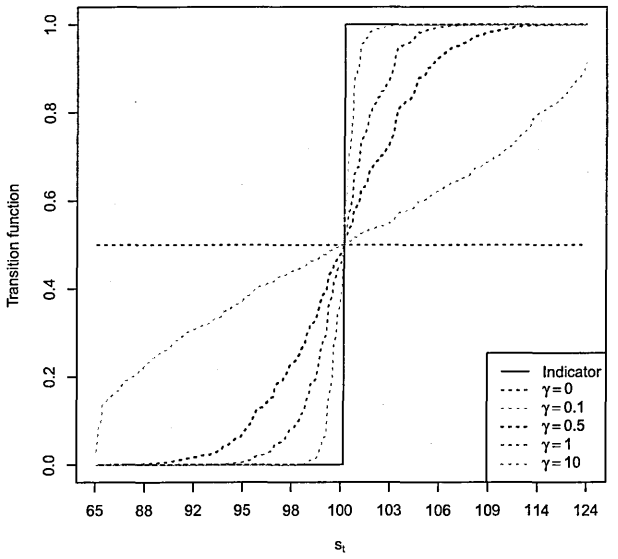


Figure 2.3: Logistic function $\pi(s_t, \gamma, c)$, with $\gamma = 0, 0.1, 0.5, 1, 10$ and $c = 100$, together with the indicator function $I(s_t; c), c = 100$, both as the transition function.

order logistic function defined as

$$\pi(s_t; \gamma, c) = [1 + \exp\{-\gamma(s_t - c)\}]^{-1} \tag{2.13}$$

The parameter γ is responsible for the degree of smoothness of the transition between the two regimes and the parameter c represents the moment when the transition occurs. Figure 2.3 illustrates the logistic function $\pi(s_t, \gamma, c)$, when $\gamma = 0, 0.1, 0.5, 1, 10$ with threshold $c = 100$, together with the indicator function $I(s_t; c)$ from equation (2.9). Notice that when γ tends to zero the logistic function becomes constant ($\pi(s_t; 0, c) = 0.5$) and the STAR model is reduced to an average between two linear AR models. As the values of γ increase, the logistic function tends to an indicator function and the transition from one regime to the other becomes more abrupt, resembling a TAR model.

A strategy for statistical parametric inference in STAR models was proposed by van Dijk *et al.* (2002) who used ordinary least squares for estimating the autoregressive coefficients and nonlinear least squares for the parameters within the logistic function. An important point to notice is that the nonlinear least squares used by classical STAR models adopt linear approximations of the nonlinear function at each iteration step. In proposed DBSTAR models, this approximation is carried out at the modelling stage rather than at the estimation stage as we shall see in chapters 4 and 5.

There have been many applications of STAR models to time series data, including the benchmark Canadian lynx (Tong, 1990).

STAR models have some limitations in their model structure. Like TAR and SETAR models, discontinuity problems in STAR models are detected when the smoothing parameter γ takes very large values (10 or larger), since the smoothing transition function then becomes an indicator transition function. In this case, all the limitations of TAR/SETAR models, that is, the requirement of large data sets, fixed parameters including the variance and stationarity, are also observed.

An additional problem is related to the convergence of the nonlinear least squares algorithm to estimate the smoothness γ and the threshold c parameters. This algorithm is an iterative procedure that requires the choice of initial values for the parameters. So different convergences may be obtained for different starting values. Another problem with the nonlinear least squares algorithm is that, despite some choices that can guarantee local convergence, global convergence can be very slow and is not guaranteed, as shown in Teräsvirta (2005). Therefore, STAR models are

not generally appropriate for real-time applications, especially in areas where the time series are observed in very short periods of time.

STAR as well as TAR/SETAR models use a high autoregressive order which tries to capture the periodic behaviour. By doing so, the models may not be parsimonious and, what is more, they may not give accurate specification of the cyclical/seasonal behaviour of the time series.

They also demand data transformation during exploratory analysis, in case the data do not meet all the requirements, such as stationarity and homoscedasticity.

2.4 Bayesian STAR models

The vast majority of nonlinear autoregressive time series models make use of classical inference. However, Bayesian approaches in this area have also been proposed recently. In this section, some Bayesian nonlinear autoregressive models are introduced and their properties are discussed. Some identified issues are also presented.

2.4.1 Bayesian TAR models

Many Bayesian TAR models and their variants have appeared in the literature. The Bayesian framework considers the model parameters, $\underline{\alpha}, c$ and σ^2, as random quantities and interpret probabilities as degrees of belief about particular realisations of each random quantity (or variable) conditional upon the availability of information.

Broemeling and Cook (1992) were the first to propose Bayesian analyses of two-regime threshold autoregressions. Given the data, Bayesian inference updates prior beliefs, expressed before observing the data, into posterior beliefs via Bayes' theorem,

as follows

$$p(\underline{\theta}|y_t) = \frac{f(y_t|\underline{\theta})f(\underline{\theta})}{f(y_t)} \tag{2.14}$$

where, y_t is the sample observation at time $t = 1, \ldots, T$, $\underline{\theta} = (\underline{\alpha}, c, \sigma^2)$ is the vector of the model parameters, $f(\underline{\theta})$ is the prior probability distribution that quantifies the beliefs before observing the data, $f(y_t|\underline{\theta})$ is the probability distribution of the sample observations conditional on the parameters (i.e., the likelihood), $f(y_t)$ is the marginal probability distribution of the sample and, finally, $p(\underline{\theta}|y_t)$ is the posterior probability distribution of the parameters that is updated after observing the data.

Chen (1998) proposed the Bayesian Generalised TAR model by incorporating exogenous variables into Bayesian TAR models and also allowing the transition variable s_t to be a function of the exogenous variables. Using notation from equations (2.8) and (2.9), a two-regime Bayesian Generalised model is defined as:

$$Y_t = \underline{z}_t'\underline{\alpha}_1 + \underline{z}_t'\underline{\alpha}_2 I(s_t; c) + \underline{X}_t'\underline{\beta} I(s_t; c) + \epsilon_t \tag{2.15}$$

where, $\underline{X}_t$ denotes a vector of set of exogenous variables in each regime and $\underline{\beta}$ is the vector of coefficients associated with the exogenous variables. Observe that the Bayesian Generalised model:

- is reduced to a Bayesian TAR model if the exogenous variables $\underline{X}_t$ are deleted.
- is reduced to a switching regression (Quandt, 1958) if the lagged variables $\underline{z}_t'$ are deleted.
- becomes a nonhomogeneous linear regression model if only the noise variance σ^2 is different for each regime.

To implement the Bayesian Generalised TAR models, the conditional posterior distribution of each parameter, given all the others, is derived. To this end, prior distributions are specified: (i) independent normal distributions for the parameters $\underline{\alpha}_1, \underline{\alpha}_2, \underline{\beta}$, (ii) independent inverse gamma distributions for the variances at each regime σ^2, (iii) a continuous uniform distribution for the threshold parameter c and (iv) a discrete uniform distribution for the delay parameter d in case the transition variable s_t is set as a lagged exogenous variable x_{t-d}. In summary, all conditional densities were identified with the exception of c, when Metropolis-Hasting algorithm (Gilks *et al.*, 1996) was employed.

In general, Bayesian parametric inference for nonlinear autoregressive models rely on computer intensive numerical simulation methods due to the loss of analytical tractability in calculating approximate posterior distributions of underlying parameters. They are, thus, not generally appropriate for real-time applications because of computation time for small time intervals, for which sequential prior-to-posterior parametric updating and forecasting would be more pertinent.

Bayesian Generalised TAR models avoid discontinuity problems in classical TAR models by using Markov Chain Monte Carlo (MCMC) algorithms to assess the posterior distributions of the parameters. In this case, they also address some of the limitations of the classical TAR models, as they do not require large data sets to estimate the parameters. However, the other limitations still stand: Bayesian Generalised TAR models have fixed parameters, and hence are not suitable for non-stationary processes, and they also demand data transformation during exploratory analysis, in case the data do not meet all the requirements.

2.4.2 Bayesian SETAR models

Bayesian versions of SETAR models have also been proposed in the literature. Geweke and Terui (1993) were the first to propose Bayesian SETAR models. Basically, the model in (2.10) is extended in the sense that probability distributions are given to its parameters.

In addition, the autoregressive order p and the delay parameter d can also be added to the vector $\underline{\theta}$ in (2.14). By doing so, the AIC criterion is no longer used to select p and d, as in classical inference, but prior probability distributions are given to p and d, and posterior probability distributions are assessed after collecting data.

Chen and Lee (1995) introduced a Bayesian analysis to the TAR model in (2.7) by transforming it into a change point problem. Intractable joint distributions were then explored via MCMC to assess the posterior distributions.

So and Chen (2003) also developed an MCMC algorithm for SETAR models, which extracts marginal posterior probability distributions for the threshold and delay parameters, c and d, respectively, from the fully conditional probability distribution for $\underline{\theta}$, which is not easily obtained.

Campbell (2004) developed a reversible jump MCMC algorithm (Green, 1995) for SETAR models with different unknown autoregressive orders p in each one of the regimes.

Koop and Potter (1999, 2003) compared classical and Bayesian SETAR models through Bayes' factors. They considered Bayes' factors as the ratio of marginal likelihoods of linear and nonlinear models. They concluded that Bayesian TAR models provide superior statistical evidence over classical tests in assessing the significance

of nonlinearities in economic time series. It suggested that the use of Bayes' factors was better at identifying nonlinearities than the hypothesis tests of classical methods.

The Bayesian formulations of SETAR models still present some limitations. Various MCMC algorithms have been proposed to overcome those and there is still some ongoing research in this area.

2.4.3 Bayesian STAR models

Bauwens *et al.* (1999) was the first work to propose computational Bayesian approaches for STAR models as given in (2.12). They used a sampling importance resampling algorithm (Gelfand and Smith, 1990) for posterior inference of parameters of a STAR model with known AR order p.

Lopes and Salazar (2005) further developed Bayesian STAR models where (i) a Gibbs sampler approach was used for inferences on the parameters $\underline{\alpha}_1, \underline{\alpha}_2, \gamma, c, d$ and σ^2 of the logistic STAR in (2.12) and (2.13) when the AR order p is known, and (ii) a reversible jump MCMC algorithm for posterior assessment of unknown p was included.

We refer to Lopes and Salazar (2005) approach as computational Bayesian STAR (CBSTAR) models to differ from Bayesian STAR models proposed in this thesis. The former, uses computer intensive numerical simulation methods due to the loss of analytical tractability in calculating approximate posterior distributions of underlying parameters. The latter, uses sequential prior-to-posterior updating analysis due to conjugacy properties.

CBSTAR models are based on Lubrano's prior distributions as in Bauwens *et al.* (1999). Conditional posteriors obtained via Gibbs sampling algorithm for the AR coefficients $\underline{\alpha}_1$ and $\underline{\alpha}_2$ follow a normal distribution with the variance σ^2 having an inverse gamma distribution. A Metropolis-Hasting algorithm was adopted to obtain the conditional posterior distributions for the nonlinear parameters γ and c. They considered $s_t = y_{t-d}$ in (2.13) and obtained a conditional posterior distribution for the delay parameter d as $p(d|y, \underline{\Theta}) \propto p(y|d, \underline{\Theta})p(d)$ from a discrete prior set for $d \in \{d_1, d_2, \ldots, d_{max}\}$, where $d_1 \leq \ldots \leq d_{max}$, for a large upper bound d_{max} with $p(d_i) = Pr(d = d_i)$.

Lopes and Salazar (2005) concentrated their attention on Bayesian STAR models where the transition function has the form of a logistic function. Their work focused on proposing a reversible jump MCMC algorithm that accounts for uncertain model order, by assuming that p is one of the parameters of the model. The model selection was based on three different information criteria: AIC, Bayesian Information Criteria (BIC) (Schwarz, 1978), and Deviance Information Criteria(DIC). The last criterion was developed and discussed in Spiegelhalter *et al.* (2002).

2.5 Variants of STAR models

2.5.1 Multiple Regimes STAR models

Even though two regimes might be sufficient in many applications, it can be desirable to obtain a STAR model that accommodates more than two regimes. An alternative

transition function in STAR models is the exponential function defined as

$$\pi(s_t;\gamma,c) = 1 - \exp\{-\gamma(s_t - c)^2\} \tag{2.16}$$

where the parameters have similar interpretations to the logistic function defined in (2.13). With this transition function, exponential STAR models present three regimes: the first regime is for observations at left upper corner of the exponential function, the second regime is for observation in the middle of the curve and the third regime for observations at the right upper corner of the exponential function.

A more general form to allow for $m = 2^k, k > 1$, different regimes was proposed by van Dijk and Franses (1999), which extends STAR models, called multiple regimes STAR models.

The multiple regimes STAR models with m regimes are defined by adding $m-1$ transition variables, as well as $m-1$ smoothness and threshold parameters to STAR models in (2.12), as

$$Y_t = \underline{z}_t'\underline{\alpha}_1 + \underline{z}_t'\underline{\alpha}_2\pi_1(s_{1t};\gamma_1,c_1) + \ldots + \underline{z}_t'\underline{\alpha}_m\pi_{m-1}(s_{(m-1)t};\gamma_{m-1},c_{m-1}) + \epsilon_t \tag{2.17}$$

where $\pi_j(s_{jt};\gamma_j,c_j), j = 1,\ldots,m-1$, are the transition functions in the range [0, 1] as in (2.13) with several transition variables s_{jt}, smoothness parameters $\gamma_j > 0$ and threshold parameters c_j. Notice that subscripts j have been added to the parameters, one for each of the m regimes. It is possible to have a single transition variable $s_t = s_{jt}$ for all regimes. In addition, when smoothness parameters become very large, STAR models become SETAR models with m regimes.

Multiple regimes STAR models have been used to describe business-cycle asymmetry (van Dijk and Franses, 1999; Dufrénot *et al.*, 2003). With STAR models (2 regimes) it is possible to detect only two states of the economy – contraction and expansion. Whereas applying multiple regimes STAR models, different phases of contractions and expansions can be explored. van Dijk and Franses (1999) applied multiple regimes STAR models with $k = 2$, hence 4 regimes, to Y_t = US real Gross National Product. A variable nominated as Current Depth Recession (CDR) to detect whether growth is increasing or decreasing was used as a transition variable $s_{1t} = CDR_{t-2}$ as well as a second transition variable to measure whether the level of real Gross National Product is above or below its previous value ($s_{2t} = \Delta y_{t-1}$), where Δy_{t-1} is the difference between y_{t-1} and y_{t-2}.

2.5.2 Time-Varying STAR models

Most of the evidence for nonlinearity has been obtained under the assumption of parameter constancy. However, there are some processes that present both nonlinearity and a structural break. The structural break is an unexpected shift in the time series.

Lundbergh *et al.* (2003) developed a STAR model considering the idea of making it possible to model both smooth transition-type nonlinearity and a structural break, simultaneously. So-called time-varying STAR models are special cases of multiple regimes STAR models. They describe Y_t by STAR models considering the time t as

the transition variable in the second transition function, as follows

$$Y_t = \underline{z}_t'\underline{\alpha}_1 + \underline{z}_t'\underline{\alpha}_2\pi_1(s_{1t};\gamma_1,c_1)\underline{z}_t'\underline{\alpha}_2\pi_2(t;\gamma_2,c_2) + \epsilon_t \tag{2.18}$$

van Dijk *et al.* (2003) show an application of this model to investigate business cycle fluctuations on seasonal patterns with a single structural break.

There are some improvements in this model compared with previous STAR models regarding changes over time in the parameters, but time-varying STAR models still present some limitations. The autoregressive coefficients change at each regime, which are determined by the second transition variable $s_{2t} = t$. However, those changes are not appropriate for modelling non-stationary processes. Although the model (2.18) is called a time-varying model, it is only suitable for processes presenting a single or few structural breaks. If there are too many unknown breaks, the underlying process is considered non-stationary and, thus, this model is no longer appropriate. In such cases, time-varying autoregressive coefficients or explicit components for modelling the (upward or downward) trend should be allowed in the model structure.

2.6 Summary

The possibility of modelling nonlinear autoregressive processes with two regimes by considering a model that can detect change and switch between regimes was well-investigated by Tong (1978). TAR models were proposed in that paper, which use an indicator function as the transition function to abruptly switch from an AR model

to another. Tong and Lim (1980) extended TAR models by treating lagged variables (Y_{t-d}) as transition variables and, thus, defining SETAR models. TAR and SETAR are piecewise linear models and can therefore have their parameters estimated by ordinary least squares. They have been well-used in many time series data applications, including some benchmark data sets, such as Canadian lynx data. A problem with those two models is that their likelihood functions present discontinuities that affect statistical inference due to the abrupt switching between regimes.

A generalisation of TAR and SETAR models incorporating a smooth transition between the two regimes via a pre-defined transition function was proposed by Chan and Tong (1986a) and called STAR models. One of the main advantages in using models with smooth transitions is the possibility of specifying a gradual transition from one regime to the other in order to avoid an abrupt switching between them. However, convergence of the algorithm to estimate the smoothness γ and the threshold c parameters depends upon given initial values and can be very slow and is not guaranteed.

STAR models also require large and stationary data sets for parameter estimation. In addition, STAR models require that parameters and observational variance be fixed over time. In an attempt to model periodic processes, STAR models may not be parsimonious or may not give accurate specification of the cyclical behaviour. Generally, the data are transformed during exploratory analysis, in case they do not meet all the requirements.

The vast majority of time series models make use of classical inference, so the parameters in the models are seen as unknown quantities which can be estimated

using well-known estimation methods. For instance, ordinary least squares for estimating the autoregressive coefficients and nonlinear least squares for the parameters within the transition function.

Nevertheless, several Bayesian approaches in nonlinear autoregressive models have been proposed recently. In general, Bayesian parametric inference for nonlinear autoregressive models rely on MCMC methods due to the loss of analytical tractability in calculating posterior distributions. They are, thus, not generally appropriate for real-time applications that require sequential prior-to-posterior parametric updating and forecasting as the convergence of the chain may be slow.

Chapter 3

Dynamic Linear Models and Mathematical background

This chapter describes Dynamic Linear Models (DLMs) and some supporting mathematical methods, which are both essential for proposing DBSTAR models in the next two chapters.

The term *dynamic* in DLMs is because the models allow time-varying parameters due to changes of the process over time. These models are described in Section 3.1 and their estimation methods in Section 3.2. Bayesian model selection approaches used in the next chapters are defined in Section 3.3.

Polynomial approximations, with a focus on Taylor series expansions, are introduced in Section 3.4, and spline functions are introduced in Section 3.5: both methods are used for approximating the transition functions from STAR models, fundamental stages for proposing DBSTAR models.

During the approximation stage, systems of polynomial equations are formed.

After the estimation stage, it will be required to find estimated values of the original parameters $(\underline{\alpha}_1, \underline{\alpha}_2, \gamma, c)$. A method for finding solutions for those is presented in Section 3.6.

3.1 Dynamic Linear Models

In the last few decades, state space models have received enormous attention, having a very wide range of applications (West *et al.*, 1985; Barbosa and Harrison, 1992; West and Harrison, 1997; Dordonnat *et al.*, 2008). In a Bayesian framework, when normality and linearity are assumed, a special case of state space models are referred to as Dynamic Linear Models (DLMs) (West and Harrison, 1997).

State space models and consequently DLMs, are sequential models where attention is focused on making inferences about the future, conditional upon existing information. They present some impressive advantages, such as, much more flexibility than static models in modelling non-stationary time series and structural breaks, and also they can be interpreted more easily. These characteristics have attracted researchers from many fields.

Given a time series $Y_t, t = 1, \ldots, T$, DLMs assume that there exists an underlying state process $\underline{\theta}_t, t = 0, \ldots, T$ that generates the observable time series, at each time. Notice that $\underline{\theta}_0$ represents the initial beliefs of the underlying process. Figure 3.1 shows a graphical representation of the conditional independence structure of a DLM. DLMs must satisfy the following two fundamental assumptions: (A) $\underline{\theta}_t$ is a Markov chain, that is, $\underline{\theta}_t$ depends only on $\underline{\theta}_{t-1}$ and a random error (not illustrated in the figure for simplicity), refer to the arrows in the figure from $\underline{\theta}_{i-1}$ to $\underline{\theta}_i, i =$

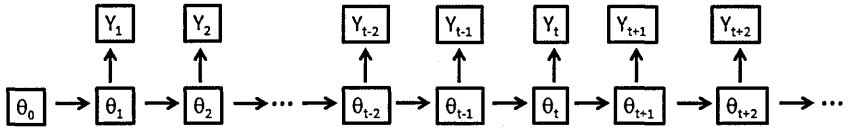


Figure 3.1: Graphical representation of the conditional independence structure of a DLM

$1, \ldots, t+3$, and (B) conditional on $\underline{\theta}_t$, the observed values of Y_t are independent and also Y_t depends only on the state process $\underline{\theta}_t$ and a random error term, refer to the arrows from $\underline{\theta}_j$ to $Y_j, j = 1, \ldots, t+2$.

West and Harrison (1997) emphasised some fundamental principles of DLMs which are summarised as follows,

- Parametrisation may be dynamic (sequentially over time)
- Information may be represented as probability
- Forecasts are produced as probability distributions
- Useful information from experts may be incorporated

The most attractive advantage of DLMs is when online inference is required. These models allow fast computations recursively, i.e., conjugate analytical models, according to the arrival of observations sequentially over time. It is a consequence of Bayes' theorem in equation (2.14) when the posterior probability distribution of the parameters $\underline{\theta}_t$ is updated, incorporating new information and assimilating past history.

The essence of the Bayesian approach is associated with probabilistic representation of all uncertain knowledge about the future (Jeffreys, 1961). The initial information, including history, is used to form initial relevant views about the future

of all defining unknown model quantities $(\underline{\theta}_t)$. As time evolves, new information may be incorporated by updating or reviewing the modeller's view of the future. The beliefs of the modeller are structured in terms of a parametric model, where the parameters are dynamic, i.e., subject to variation over time to adapt influential information affects.

Let $D_t = (y_t, D_{t-1})$ represent the information available at time t after observing Y_t. So, D_0 represents the existing information available to, and recognised by, a modeller at time $t = 0$. Similarly, at any time t, $t > 0$, the existing information is represented by D_t with the difference that D_t includes D_0 and the values of the observations $y_t, y_{t-1}, ..., y_1$. In summary, for $t > 0$, D_t comprises all the information available at time t. Thus, the only new information becoming available is the observed value y_t.

The basic formulation of DLMs is specified by the following three probabilities:

the conditional distribution of the observation,

$$(Y_t \mid \underline{\theta}_t) \sim N\left(\underline{F}_t'\underline{\theta}_t, V_t\right) \tag{3.1}$$

the conditional distribution of the parameters (state),

$$\left(\underline{\theta}_t \mid \underline{\theta}_{t-1}\right) \sim N\left(\mathbf{G}_t\underline{\theta}_{t-1}, \mathbf{W}_t\right). \tag{3.2}$$

and the initial information of the parameters,

$$\left(\underline{\theta}_0 \mid D_0\right) \sim N\left(\underline{m}_0, \mathbf{C}_0\right), \tag{3.3}$$

DLMs are completely determined by initial information with probability distribution $p(\underline{\theta}_0 \mid D_0)$ and the conditional probability densities $p(Y_t|\underline{\theta}_t)$ and $p(\underline{\theta}_t|\underline{\theta}_{t-1})$.

For each time t, define a quadruple $\Delta_t = \{\underline{F}_t, \mathbf{G}_t, V_t, \mathbf{W}_t\}$, where $\underline{F}_t$ is a design vector of known values of variables, $\mathbf{G}_t$ is an evolution matrix, V_t is the unknown scalar observational variance and $\mathbf{W}_t$ is an unknown evolution covariance matrix. A DLM is then specified by the quadruple Δ_t.

DLMs have also the advantage of incorporating components into the quadruple Δ_t, representing different characteristics of the underlying process, such as trend, cycles and seasonality.

3.1.1 AR models in DLM form

Any AR model by Box and Jenkins (1970) can be written in DLM form. More formally, it is possible to find a DLM whose measurement process (Y_t) has the same distribution as the given AR model.

Many different forms of representing AR models have been proposed in the literature. In this thesis, only one form is considered, which is probably the most widely used – see West and Harrison (1997, p.272,305) for further discussion.

Example 3.1. *Consider the linear AR(p) models defined by equation (1.2). Consider also the quadruple Δ_t that defined DLMs. For an autoregressive order $p = 1$, AR(1) models can be rewritten into DLMs form with $\underline{F}_t' = [1, Y_{t-1}]$, the 2×2-dimensional $\mathbf{G}_t = \mathbf{I}$ (the identity matrix), the scalar $V_t = \sigma_\epsilon^2$ and a null 2×2 state variance $\mathbf{W}_t = \mathbf{0}$, and a state vector $\underline{\theta}_t = [\theta_{0t}, \theta_{1t}]$ associated with $\underline{F}_t'$. Consider now an AR(2) model. Then, the components of the quadruple Δ_t are expanded to*

$\underline{F}_t' = [1, Y_{t-1}, Y_{t-2}]$, 3×3 *identity matrix* $\mathbf{G}_t = \mathbf{I}$, $V_t = \sigma_\epsilon^2$ *and* $\mathbf{W}_t = \mathbf{0}$, *as well as* $\underline{\theta}_t = [\theta_{0t}, \theta_{1t}, \theta_{2t}]$.

From Example 3.1, it becomes straightforward that AR(p) models can be written in DLM form by defining the quadruple Δ_t to be: $\underline{F}_t' = [1, Y_{t-1}, Y_{t-2}, ..., Y_{t-p}]$, the $(p+1) \times (p+1)$-dimensional identity matrix $\mathbf{G}_t = \mathbf{I}$, $V_t = \sigma_\epsilon^2$ and the null $(p+1) \times (p+1)$-dimensional state matrix $\mathbf{W}_t = \mathbf{0}$. Notice that, AR(p) models are for modelling stationary processes, therefore the state matrix $\mathbf{W}_t$ has zero elements in all positions, which gives all the parameters the aspect of being fixed over time, i.e., $\underline{\theta}_{t-1} = \underline{\theta}_t$, for all t.

3.2 Parametric prior-to-posterior updating

In this section, a parametric prior-to-posterior updating procedure to linear filtering, smoothing and forecasting problems adapted from Kalman (1960) is described. It was originally related to physical systems in engineering applications, nevertheless, due to computation simplicity, easy implementation and fast parameter estimation, the Kalman filter, as it is known, has been extensively applied in different areas, including time series analysis.

On the one hand, this parametric prior-to-posterior updating procedure sequentially over time uses past data to make inference about parameters incorporating information collected at the same time. This stage is referred to as filtering. On the other hand, the use of recent data to revise the filtered information, i.e., estimate the parameters retrospectively backward in time, is referred to as smoothing. Moreover, the past and the current data are used to make inference about the future, for which

it is referred to as forecasting.

Consider a DLM with a constant but unknown observation variance, i.e., $V_t = V$. The quadruple changes to $\Delta_t = \{\underline{F}_t, \mathbf{G}_t, V, V\mathbf{W}_t\}$. Notice that the system variance matrix $\mathbf{W}_t$ is scaled by the unknown observational variance V. When Δ_t is defined in this way, a Bayesian conjugate prior-to-posterior analysis can be developed by assuming that $(\underline{\theta}_t, V|D_t)$ has a multivariate normal/inverse gamma distribution (see Gelman *et al.* (1995)). The marginal posterior distribution of interest $(\underline{\theta}_t|D_t)$ has a multivariate Student-t distribution and $(V|D_t)$ an inverse gamma distribution.

The Kalman filter is an algorithm that calculates the posterior distribution for $\underline{\theta}_t$ sequentially over time, given all information D_t, $(\underline{\theta}_t|D_t)$, due to parametric prior-to-posterior conjugacy above mentioned. It is repeated each time that a new observation y_t becomes available. Also, with the prior distribution for $\underline{\theta}_t$ we can calculate the 1-step ahead forecast distribution of Y_t. West and Harrison (1997) referred to this procedure as the Bayesian forecasting system, which can be used when both variances V and $\mathbf{W}_t$ are assumed to be either known or unknown.

3.2.1 Prior distributions

The initial information at time $t = 0$, including history, is used to form initial relevant views about the future for all model parameters. In case of no prior knowledge about the parameters, non-informative prior distributions can be used. At time $t = 0$, the distribution of the state vector is given by the specification of a

multivariate Student-t distribution with n_0 degrees of freedom

$$(\underline{\theta}_0 \mid D_0) \sim T_{n_0}(\underline{m}_0, \mathbf{C}_0)$$

where, $\underline{m}_0$ and $\mathbf{C}_0$ are the prior mean vector and covariance matrix. The initial distribution for the observational variance V_0 is an inverse gamma with hyperparameters n_0 and S_0

$$(V_0 \mid D_0) \sim IG\left(\frac{n_0}{2}, \frac{n_0 S_0}{2}\right).$$

3.2.2 Filtering distributions

The parametric prior-to-posterior updating procedure sequentially over time uses past data to make inference about parameters incorporating information collected at same time. At time $t-1$, all beliefs about the parameters θ_{t-1} are represented by the posterior probability distribution

$$(\underline{\theta}_{t-1} \mid D_{t-1}) \sim T_{n_{t-1}}(\underline{m}_{t-1}, \mathbf{C}_{t-1})$$

where n_{t-1} is degree of freedom of a multivariate Student-t distribution, $\underline{m}_{t-1}$ and $\mathbf{C}_{t-1}$ are the posterior mean vector and covariance matrix, respectively. And also, the beliefs about the observational variance are represented by the posterior probability distribution

$$(V \mid D_{t-1}) \sim IG\left(\frac{n_{t-1}}{2}, \frac{n_{t-1} S_{t-1}}{2}\right)$$

where S_{t-1} is the point estimate of the observational variance.

At time t before observing Y_t, the posterior distribution dynamically evolves

into prior distributions. These priors distributions represent the prior beliefs of the modeller about the parameters. The prior probability distribution for the state vector $\underline{\theta}_t$ is

$$(\underline{\theta}_t \mid D_{t-1}) \sim T_{n_{t-1}}(\underline{a}_t, S_{t-1}\mathbf{R}_t)$$

where $\underline{a}_t = \mathbf{G}_t\underline{m}_{t-1}$ and $\mathbf{R}_t = \mathbf{G}_t\mathbf{C}_{t-1}\mathbf{G}_t' + \mathbf{W}_t$ are the prior mean and prior variance of $(\underline{\theta}_t \mid D_{t-1})$, respectively. And also, the prior probability distribution for the observational variance V is

$$(V \mid D_{t-1}) \sim IG\left(\frac{n_{t-1}}{2}, \frac{n_{t-1}S_{t-1}}{2}\right).$$

After observing $Y_t = y_t$, new information are incorporated into previous prior probabilities and the modeller's beliefs are updated, being then represented by posterior distributions. The posterior probability distribution for the state vector $\underline{\theta}_t$

$$(\underline{\theta}_t \mid D_t) \sim T_{n_t}(\underline{m}_t, \mathbf{C}_t)$$

where $n_t = n_{t-1} + 1$, $\underline{m}_t = \underline{a}_t + (\mathbf{R}_t\underline{F}_t/Q_t)e_t$ and $\mathbf{C}_t = \frac{S_t}{S_{t-1}}(\mathbf{R}_t - \mathbf{R}_t\underline{F}_t/Q_t\underline{F}_t'\mathbf{R}_t')$, and the posterior probability distribution for the observational variance V

$$(V \mid D_t) \sim IG\left(\frac{n_t}{2}, \frac{n_tS_t}{2}\right)$$

with $S_t = S_{t-1} + \frac{S_{t-1}}{n_t}(\frac{e_t^2}{Q_t} - 1)$ and $e_t = y_t - f_t$, the one-step-ahead forecasting error.

This sequential algorithm repeats these steps over time incorporating new information and assimilating past history.

Consider now that the unknown observational variance changes over time (V_t). It is desirable to retain the gamma form for the resulting prior distribution of $(V_t \mid D_{t-1})$ as well as the posterior distribution of $(V_t \mid D_t)$ so the conjugacy analysis models can be straightforwardly adapted. In such case, it is possible to use an appropriately chosen variance discounting technique for V_t, which is assumed to change but only slowly and steadily over time to avoid potential unpredictable behaviour that can lead to loss of analytical tractability (Broemeling, 1985). Therefore, for an unknown observation covariance V_t, this prior-to-posterior approach can be adapted for which a discount factor δ_V, satisfying the condition $0 < \delta_V \leq 1$, is considered.

Then, at time $t-1$, the observational variance has the posterior probability distribution

$$(V_{t-1} \mid D_{t-1}) \sim IG\left(\frac{\delta_V n_{t-1}}{2}, \frac{\delta_V n_{t-1} S_{t-1}}{2}\right)$$

At time t before observing Y_t, the posterior distribution dynamically evolves into prior distribution as

$$(V_t \mid D_{t-1}) \sim IG\left(\frac{\delta_V n_{t-1}}{2}, \frac{\delta_V n_{t-1} S_{t-1}}{2}\right)$$

Comparing the expected value of prior and posterior distributions of the observation variance, respectively, it is possible to see that $E(V_t \mid D_{t-1}) = E[V_t \mid D_t] = \frac{1}{S_{t-1}}$, that means that both inverse gamma distributions have the same location. However, the dispersion increased with the inclusion of the discounting of the degrees of freedom parameter, i.e., $\delta_V n_{t-1} < n_{t-1}$. For further discussion, see West and Harrison (1997, p.111,360-362) and references therein.

At time t after observing $Y_t = y_t$, the posterior distribution for the observational variance is updated to

$$(V_t \mid D_t) \sim IG\left(\frac{\delta_V n_t}{2}, \frac{\delta_V n_t S_t}{2}\right)$$

To update this distribution, the observational variance discount factor δ_V has to be included into the hyperparameters of the gamma distribution. Thus, $n_t = \delta_V n_{t-1} + 1$ and $S_t = S_{t-1} + \frac{S_{t-1}}{\delta_V n_t}(\frac{e_t^2}{Q_t} - 1)$ in the usual notation.

Furthermore, the unknown state covariance matrix $\mathbf{W}_t$ may change over time. For similar reasons regarding conjugacy, it is also possible to use an appropriately chosen variance discounting technique for $\mathbf{W}_t$ assuming that the parameters $\underline{\theta}_t$ change but only slowly and steadily over time to avoid loss of analytical tractability (Broemeling, 1985).

Therefore, this prior-to-posterior approach can be adapted for which a discount factor δ_W, satisfying the condition $0 < \delta_W \leq 1$, is considered as follows,

$$\mathbf{W}_t = \left(\frac{1 - \delta_W}{\delta_W}\right) \mathbf{G}_t \mathbf{C}_{t-1} \mathbf{G}_t'.$$

where $\mathbf{C}_{t-1}$ is a prior covariance matrix for $\underline{\theta}_t$. Notice that when $\delta_W = 1$, the state variance is a null matrix $\mathbf{W}_t = \mathbf{0}$ which means that the state process is constant over time, whilst the smaller the δ_W, the larger the values of $\mathbf{W}_t$ and consequently more variability is present in the state vector $\underline{\theta}_t$.

Usually, δ_V takes values between 0.9 and 0.99 and δ_W takes values between 0.8 and 0.99 (West and Harrison, 1997).

3.2.3 Smoothing distributions

Recent data are used to revise the filtered distribution of the parameters in order to understand better what happened to the estimates parameters retrospectively backward in time. This procedure can be implemented with the use of the Kalman smoother approach which can be straightforwardly implemented, like the Kalman filter, as a backward-recursive algorithm. This algorithm depends only on the data used for filtering and the one-step-ahead forecast moments. A reason for using the Kalman smoother algorithm after filtering is that it makes static and dynamic models comparable in the sense that it uses the whole data set for estimating backward the parameters. West and Harrison (1997) refer to this procedure as retrospective analysis.

Firstly, we need the posterior probability distribution for $\underline{\theta}_t$, at each time t, provided by the Kalman filter. Thus, we go (forward) filtering $(\underline{\theta}_t|D_t)$. The backward-recursive algorithm starts with

$$(\underline{\theta}_T \mid D_T) \sim T_{n_t}(\underline{a}_t(0), \mathbf{R}_t(0))$$

where $\underline{a}_t(0) = \underline{m}_T$, the posterior mean vector and $\mathbf{R}_t(0) = \mathbf{C}_T$, the posterior covariance matrix, both computed by the last updating of the Kalman filter. Then, we go (backward) smoothing this distribution to provide the smoothing probability distribution for $(\underline{\theta}_t|D_T)$, for $t = T-1, T-2, \ldots, 1$. Therefore, the Kalman smoother algorithm provides the conditional probability distributions of $\underline{\theta}_t$ given the data D_T, for any time $t < T$.

For $1 \leq k \leq t$, the smoothing probability distributions for the state are

$$(\underline{\theta}_{t-k}|D_T) \sim T_{n_t}\left[\underline{a}_t(-k), \frac{S_t}{S_{t-k}}\mathbf{R}_t(-k)\right]$$

where $\underline{a}_t(-k) = \underline{m}_{t-k} + \mathbf{C}_{t-k}\mathbf{G}'_{t-k+1}\mathbf{R}^{-1}_{t-k+1}(\underline{a}_t(-k+1) - \underline{a}_{t-k+1})$ and $\mathbf{R}_t(-k) = \mathbf{C}_{t-k} - \mathbf{C}_{t-k}\mathbf{G}'_{t-k+1}\mathbf{R}^{-1}_{t-k+1}(\mathbf{R}_{t-k+1} - \mathbf{R}_t(-k+1))\mathbf{R}^{-1}_{t-k+1}\mathbf{G}_{t-k+1}\mathbf{C}'_{t-k}$. And the corresponding smoothed probability distribution for the level of the series are

$$(\underline{\mu}_{t-k}|D_T) \sim T_{n_t}\left[f_t(-k), \frac{S_t}{S_{t-k}}\underline{F}'_{t-k}\mathbf{R}_t(-k)\underline{F}_{t-k}\right]$$

where $f_t(-k) = \underline{F}'_{t-k}\underline{a}_t(-k)$.

3.2.4 Forecast Distributions

The past and the current data are used to make inference about the future, for which it is referred to as forecasting. Conditional forecast distributions for the states and observations at a future time $t + k$, given the data up to time t, are recursively determined. Since these forecast distributions are Student-t, it is enough to compute their means and variances.

The 1-step ahead forecast probability distribution for y_t is

$$(Y_t \mid D_{t-1}) \sim T_{n_{t-1}}(f_t, Q_t)$$

with $f_t = \underline{F}'_t\underline{a}_t$ and $Q_t = \underline{F}'_t\mathbf{R}_t\underline{F}_t + S_{t-1}$.

With $\underline{a}_t(0) = \underline{m}_t$ and $\mathbf{R}_t(0) = \mathbf{C}_t$, i.e., the distributions for the future take

information from the filtering distributions at time t, the $k \geq 1$ -step ahead forecast probability distribution for the state vector is

$$(\underline{\theta}_{t+k} \mid D_t) \sim T_{n_t}\left(\underline{a}_t(k), S_t \mathbf{R}_t(k)\right)$$

where $\underline{a}_t(k) = \mathbf{G}_{t+k}\underline{a}_t(k-1)$ and $\mathbf{R}_t(k) = \mathbf{G}_{t+k}\mathbf{R}_t(k-1)\mathbf{G}'_{t+k} + \mathbf{W}_{t+k}$. And for the future observation y_{t+k}, the forecast probability distribution is

$$(Y_{t+k} \mid D_t) \sim T_{n_t}\left(f_t(k), Q_t(k)\right)$$

with $f_t(k) = \underline{F}'_{t+k}\underline{a}_t(k)$ and $Q_t(k) = \underline{F}'_{t+k}\mathbf{R}_t(k)\underline{F}_{t+k} + S_{t+k}$.

3.3 Bayesian model selection for determination of parameters

In this section, we describe a Bayesian approach to obtain values of unknown quantities, such as autoregressive order p and delay parameter d, which cannot be accommodated in the state vector $\underline{\theta}_t$ of DLMs. This method is based on either fitting, smoothing or predictive performance of different models and involves sequentially processing observations $y_1, y_2, \ldots, y_t$, easily calculated from the one-step ahead predictive distribution $p(y_t \mid D_{t-1})$. Notice that it is important to keep the same model structure to use this method, differing only in the values of parameters. For a philosophical viewpoint, see Good (1985); for a general discussion, see West (1986) and West and Harrison (1997).

3.3.1 Bayes factor criterion

Let τ represent an unknown discrete parameter, e.g., the autoregressive order p, that can take integer values for which the most likely value τ^* is to be determined. Using a Bayesian parametric prior-to-posterior updating approach, we have for each observation y_t the one-step ahead predictive density $p(y_t \mid D_{t-1}, \tau)$. It is worth clarifying that each $p(y_t \mid D_{t-1}, \tau)$ is obtained after observing $y_1, y_2, \ldots, y_t$ but before observing $y_{t+1}, y_{t+2}, \ldots, y_T$. The joint predictive density

$$p(y_t, y_{t-1}, \ldots, y_1 \mid D_0, \tau) = \prod_{t=1}^{T} p(y_t \mid D_{t-1}, \tau) \tag{3.4}$$

is the predictive likelihood of all observations $y_t, y_{t-1}, \ldots, y_1$ conditional on the unknown τ. We use the Bayes' factor (Jeffreys, 1961), or weights of evidence (Good, 1985), defined as the ratio between the predictive likelihoods of two models, say model A and model B, differing only in values of the parameter τ, evaluated at the observed values of $Y_t = y_t, t = 1, \ldots, T$. Thus, from (3.4), the logarithm of the Bayes' factor is

$$\log(BF) = \sum_{t=1}^{T} \log\left[p_A(y_t \mid D_{t-1}, \tau = \tau_1)\right] - \sum_{t=1}^{T} \log\left[p_B(y_t \mid D_{t-1}, \tau = \tau_2)\right] \tag{3.5}$$

This measure gives us evidence in favour of ($\log(BF) > 0$) or against ($\log(BF) < 0$) model A relative to model B, according to their predictive performances when τ assumes either τ_1 or τ_2. The farther from zero $\log(BF)$ is, the stronger the evidence and for $\log(BF) = 0$ there is no evidence either way.

3.3.2 Log-predictive likelihood criterion

Alternatively, the idea of the Bayes' factor for choosing values for τ can be generalised when comparing more than two models. The evidence in favour of a model can also be obtained by checking which log-predictive likelihood (LPL), the log of the joint predictive density defined in (3.4), is the largest. Operationally, τ^* is chosen by calculating the LPL for different models differing only in the value of τ, and selecting the model which gives the largest LPL. Notice that, contrary to a fully Bayesian approach where all unknown quantities are included in the model parameter set $(\underline{\theta}_t)$ and in principle requires no data initially, the LPL approach requires an initial data set.

3.3.3 Log-smoothing likelihood criterion

The log-smoothing likelihood (LSL), conditional on the parameters, is defined in a similar way to the LPL from previous Section 3.3.2. The LSL gives evidence in favour of a model for fitting purpose, which presents the largest value of the joint smoothing density of all observations $y_t, y_{t-1}, \ldots, y_1$, conditional on the unknown τ, after having observed Y_T, i.e., the observation at the last time point T, as follows

$$p(y_t, y_{t-1} \ldots y_1 \mid D_T, \tau) = \prod_{t=1}^{T} p(y_t \mid D_T, \tau) \tag{3.6}$$

The LSL criterion is used alternatively to the LPL when the aim is to select dynamic models based on fitting performance rather than forecasting performance. Generally, the interest lies in improving understanding about the process in order to detect

unexplained changes. In addition, the LSL criterion can also be used for making dynamic and static models comparable.

3.4 Polynomial approximations

In this section, the use of polynomial approximations is described as simple mathematical functions generally used to represent curves which are difficult to evaluate. Our interest lies in using them to determine the underlying functional forms of non-linear transition functions, as we will see in chapters 4 and 5. A survey in polynomial approximations is given by Atkinson (1988). Notice that, we call polynomial order as n and degree as $r = n - 1$.

A general expression for polynomials $P_r(s_t)$ of degree r on a variable s_t is denoted as a weighted sum of basis functions $B_i(s_t), i = 0, \ldots, r$, and can be written in the form

$$P_r(s_t) = \sum_{i=0}^{r} \alpha_i B_i(s_t) \tag{3.7}$$

where $\alpha_i, i = 0, \ldots, r$, are coefficients of the polynomials which determine the shape of the polynomial approximation. There are many different types of polynomials in the literature. They are determined by choosing different basis functions, $B_i(s_t)$, such as power functions or other forms as shown in subsequent sections of this chapter, where three different ways to obtain approximations are described. The Weierstrass Theorem (Atkinson, 1988) affirms that there is always a polynomial uniformly suitable to approximate any continuous function defined on closed interval $[a, b]$.

A first-degree polynomial is just the equation of a straight line with a slope which characterises a linear approximation. The basis functions of this polynomial are $B_0(s_t) = 1$ and $B_1(s_t) = s_t$ and can be represented in a matrix form called a design matrix, as follows

$$B(s_t) = \begin{bmatrix} 1 & s_1 \\ \vdots & \vdots \\ 1 & s_T \end{bmatrix}$$

Polynomials with higher degrees have superiority in representing general nonlinear relationships due to their flexible approximation, such as a second-degree polynomial that describes a parabola and a third-degree polynomial representing an S-shape curve, and so forth. The flexibility comes with the increment of the polynomial degree by adding more basis functions to the design matrix. The general polynomial $P_r(s_t)$ in equation (3.7) with $B_i(s_t)$ as power functions has the design matrix as

$$B(s_t) = \begin{bmatrix} 1 & s_1 & s_1^2 & \dots & s_1^r \\ 1 & s_2 & s_2^2 & \dots & s_2^r \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & s_T & s_T^2 & \dots & s_T^r \end{bmatrix}$$

A good basis function should be chosen such that characterisation of the underlying nonlinear transition function can be well specified. Also, the polynomial approximations should be flexible enough to exhibit the required curvature where needed.

Taylor polynomials

Taylor polynomials are approximation methods used to represent underlying functions. The method requires the functions to be infinitely differentiable at the real points where the approximations are performed. The resulting polynomials are infinite sums of power series.

A Taylor polynomial $P_r(s_t)$ of degree r for a function $f(s_t)$ with a variable s_t near a value a can be given by

$$P_r(s_t) = f(a) + \frac{f'(a)}{1!}(s_t - a) + \frac{f''(a)}{2!}(s_t - a)^2 + \ldots + \frac{f^{(r)}(a)}{r!}(s_t - r)^r + R_r(s_t) \quad (3.8)$$

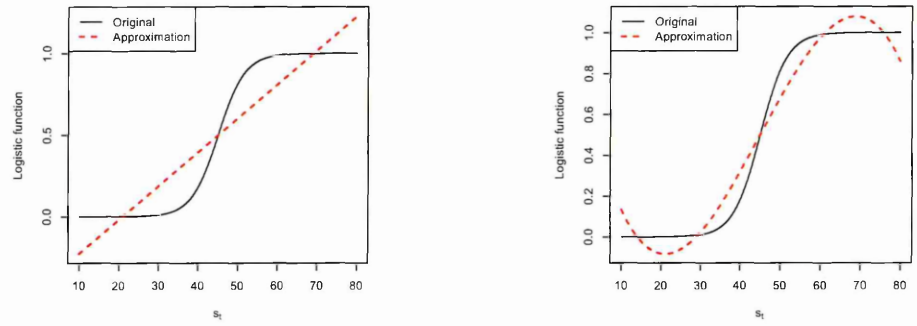
where $f^{(i)}(a), i = 1, 2, \ldots, r$, is the i-th derivative of $f(s_t)$ evaluated at the point a and $i!$ is the factorial of i. The difference between the polynomial approximation and the function is called the remainder term (Atkinson, 1988), expressed as $R_r(s_t) = P_r(s_t) - f(s_t)$. The approximation is expected to be as close to the function as possible so that $\lim_{r\to\infty} R_r(s_t) = 0$.

The function $f(s_t)$ is expanded around a given specific point a and, usually, the function is well approximated if sufficient terms are included in $P_r(s_t)$. Differentiation of power series can be performed term by term so the polynomial $P_r(s_t)$ is straightforwardly obtained.

Example 3.2. *Consider the logistic function in equation (2.13) stated again*

$$\pi(s_t; \gamma, c) = [1 + \exp\{-\gamma(s_t - c)\}]^{-1}$$

The Taylor series approximation of degree $r = 1$ of the logistic function around the



(a) (b)

Figure 3.2: Taylor series expansion of (a) degree r = 1 and (b) degree r = 3, of a logistic function around at $s_t = c$ with $\gamma = 0.3$ and $c = \bar{s}_t$

point where the transition variable assume the value $s_t = c$ *is given by,*

$$P_1(s_t) = \frac{1}{48}(24 - 12\gamma c + \gamma^3 c^3) + \frac{1}{48}(12\gamma - 3\gamma^3 c^2)s_t + R_1(s_t) \tag{3.9}$$

Similarly, the Taylor series approximation of degree $r = 3$ *of the same function at around the same point* $s_t = c$ *is given by,*

$$P_3(s_t) = \frac{1}{48}(24 - 12\gamma c + \gamma^3 c^3) + \frac{1}{48}(12\gamma - 3\gamma^3 c^2)s_t + \frac{3}{48}\gamma^3 c s_t^2 - \frac{\gamma^3}{48}s_t^3 + R_3(s_t) \tag{3.10}$$

Figure 3.2 illustrates examples of a Taylor series expansion of the logistic function from equation (2.13), at the point which the transition variable assumes the value $s_t = c$. Panel (a) illustrates the case with degree $r = 1$ from equation (3.9) and Panel (b) with degree $r = 3$ from equation (3.10). In both cases, it was assumed $\gamma = 0.3$ and $c = \bar{s}_t$. The degree $r = 1$ is definitely not recommended for this case as the straight line cannot represent the logistic function. Including more terms in the Taylor series, such as increasing the degree to $r = 3$, it approximates better

in the vicinities of $s_t = c$, whilst discrepancies are still observed at the extremes of the curve. Overall, the approximation in Panel (b) can be considered a good approximation as it maintains the S-shape of the underlying logistic function. It is an acceptable approximation but still not the most appropriate.

There may be some nonlinear transition functions that present some complexities, such as discontinuities. In such cases, the Taylor polynomials cannot represent such functions and other approximation methods, such as splines, should be used instead.

3.5 Spline functions

Taylor polynomials assume that the functions are known explicitly, so that approximations can be found based on their derivatives. However, very frequently it is necessary to obtain approximations based on very little information about the underlying process. Spline functions address situations where no fixed function is known.

Generally, polynomial approximations oscillate around a specific function value. Splines, however, represent the function by using a combination of low degree local polynomials joined together to form the polynomial $P_r(s_t)$. Also called piecewise polynomial functions, they can be determined by dividing the domain, say $[a, b]$, of the variable s_t into intervals. These intervals are separated by points that are called knots, represented by κ, which give $[a, \kappa_1, \ldots, \kappa_k, b]$. Therefore, polynomial curves are specified in each interval, $[a, \kappa_1], [\kappa_1, \kappa_2], \ldots, [\kappa_k, b]$.

In this section, two approaches are described, the Truncated Power Basis (TPB)

and B-splines. The spline polynomial $P_r(s_t)$ is obtained by imposing continuity and differentiability to the piecewise polynomials up to a certain order at the knots, where two adjacent segments join.

3.5.1 Truncated Power Basis (TPB)

To impose continuity on the piecewise polynomials at the knot κ, truncated power functions of degree r can be used. Constraints on the parameters are imposed such that the function $f(s_t)$ can be continuously approximated by the polynomial $P_r(s_t)$. Truncated power functions are defined as

$$(s_t - \kappa_i)_+^r = (s_t - \kappa_i)^r I_{s_t > \kappa_i}(s_t), i = 1, \ldots, k \tag{3.11}$$

where, $I_{s_t>\kappa_i}(s_t)$ is an indicator function and the symbol "+" means that the function takes the value 0 for s_t located at the left of κ_i and $(s_t - \kappa_i)^r$ otherwise, that is

$$(s_t - \kappa_i)_+^r = \begin{cases} 0, & \text{if } s_t \leq \kappa_i \\ (s_t - \kappa_i)^r, & \text{otherwise} \end{cases}$$

Piecewise polynomials of order n and, consequently, degree $r = n-1$ connected at knots $\kappa_1, \kappa_2, \ldots, \kappa_k$ have truncated power basis $1, s_t, \ldots, s_t^r, (s_t - \kappa_1)_+^r, \ldots, (s_t - \kappa_k)_+^r$ and their linear combination gives a spline function which has continuous derivatives up to degree r as:

$$P_r(s_t) = \sum_{a=0}^{r} \beta_a s_t^a + \sum_{b=1}^{k} \beta_{rb}(s_t - \kappa_b)_+^r \tag{3.12}$$

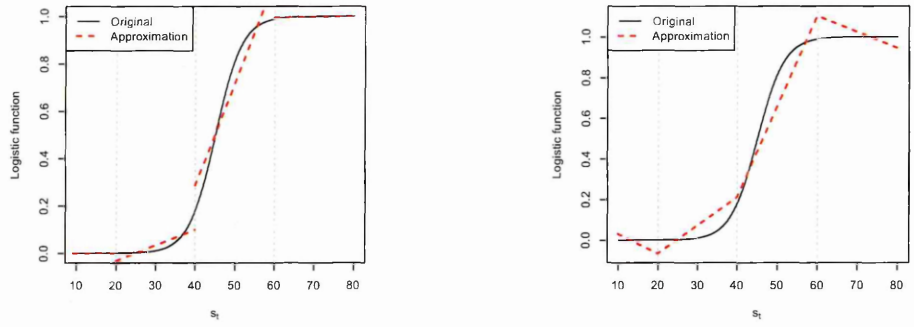
where $\beta_a, a = 0, \ldots r$ are coefficients associated with the first part of $P_r(s_t)$, for

which a polynomial of degree r, $\sum_{a=0}^{r} \beta_a s_t^a$, is specified, and $\beta_{rb}, b = 1, \ldots, k$ are coefficients associated with the second part of $P_r(s_t)$, for which a truncated power function of degree r related to the knots, $\sum_{b=1}^{k} \beta_{rb}(s_t - \kappa_b)_+^r$, is specified. Notice that the first part has sum of power series, therefore, this polynomial form is similar to that of Taylor series approximations. This characterises continuity of $P_r(s_t)$ for the whole polynomial. On the other hand, the second part is the part responsible for specifying piecewise polynomials at each interval of knots.

From the polynomials in (3.12), it is possible to extract the matrix that contains the functions of the variable s_t. That matrix is referred to as the design matrix and is specified as,

$$B(s_t) = \begin{bmatrix} 1 & s_1 & \ldots & s_1^r & (s_1 - \kappa_1)_+^r & \ldots & (s_1 - \kappa_k)_+^r \\ 1 & s_2 & \ldots & s_2^r & (s_2 - \kappa_1)_+^r & \ldots & (s_2 - \kappa_k)_+^r \\ \vdots & \vdots & \ldots & \vdots & \vdots & \ldots & \vdots \\ 1 & s_T & \ldots & s_T^r & (s_T - \kappa_1)_+^r & \ldots & (s_T - \kappa_k)_+^r \end{bmatrix} \tag{3.13}$$

Example 3.3. *Figure 3.3 illustrates two scenarios of a TPB spline function of degree $r = 1$ approximating the logistic function in equation (2.13) with variable $s_t, \gamma = 0.3$ and $c = \bar{s}_t$. Panel (a) shows the case when only the truncated power functions related to the knots is considered, that is, the piecewise polynomials $\beta_0 + \beta_1 s_t$, using only the first two columns of the design matrix (3.13). Notice that in this illustration, there are three knots, $\kappa_1 = 20$, $\kappa_2 = 40$ and $\kappa_3 = 60$, identified in the figure as dashed vertical lines. Therefore, four polynomial straight lines are specified in each interval [0, 20], [20, 40], [40, 60] and [60, 80] to approximate the logistic*



(a) (b)

Figure 3.3: (a) Piecewise polynomials of degree $r = 1$ and (b) Truncated Power Basis approximation of degree $r = 1$, of a logistic function at variable $s_t, \gamma = 0.3$ and $c = \bar{s}_t$

function. Panel (b) shows the complete TPB approximation of degree $r = 1$ *of the logistic function, that is, the piecewise polynomials* $\beta_0 + \beta_1 s_t + \sum_{b=1}^{3} \beta_{1b}(s_t - \kappa_b)_+$, *using the corresponding columns of the design matrix (3.13) when* $r = 1$ *and* $k = 3$.

The advantage of the TPB spline function is that the representation of a function is straightforwardly obtained and very attractive for statistical work. However, there are some limitations of that method, mainly associated with numerical problems, such as, rounding errors (Ramsay *et al.*, 2009), numerical instability if the number of knots is large (Ruppert *et al.*, 2003) and the design matrix may contain only a few zeros that prevents the TPB spline function from using sparse matrix techniques to reduce computational time (de Boor, 1978).

3.5.2 B-Splines

Although derived from truncated power functions, B-spline functions are piecewise polynomials with more stable numerical properties than TPB functions (de Boor,

1978).

The B-spline function extends the TPB spline function by adding n knots in the interval $[a, b]$, in a non-decreasing sequence, such as $\kappa_1 = \kappa_2 = \ldots = \kappa_{n-1} = \kappa_n < \kappa_{n+1} < \ldots < \kappa_{n+k} < \kappa_{n+k+1} = \kappa_{n+k+2} = \ldots = \kappa_{2n+k}$, where $\kappa_n = a$ and $\kappa_{n+k+1} = b$. Usually, the extra knots at the lower extreme $\kappa_1, \kappa_2, \ldots, \kappa_{n-1}$ are equal to a and at the upper extreme $\kappa_{n+k+1}, \kappa_{n+k+2}, \ldots, \kappa_{2n+k}$ are equal to b, respectively, so the sequence of knots in the interval $[a, b]$ becomes $a = \ldots = a < \kappa_{n+1} < \ldots < \kappa_{n+k} < b = \ldots = b$. In this way, the k interior knots $(\kappa_{n+1}, \ldots, \kappa_{n+k})$ in the interval $[a, b]$, excluding the extremes a and b, are similar to those knots in TPB spline function. However, it is also possible to give other arbitrary values to the extra knots. Either ways, the B-spline approach requires those knots extra to be added to the k interior knots, so the basis function can be completely specified.

B-spline approximations of order n has $m = n + k$ basis functions in their formulations, i.e., the number of columns in the design matrix is equal to the B-spline order n plus the number of interior knots k, and their linear combination gives a spline function as:

$$P_r(s_t) = \sum_{i=1}^{m} \alpha_i B_i(s_t) \tag{3.14}$$

where $\alpha_i, i = 1, \ldots, m$, are coefficients associated with the B-spline basis function $B_i(s_t)$. The basis functions $B_i(s_t), i = 1, \ldots, m$, are obtained recursively by the Cox-de Boor algorithm (de Boor, 1978), given the order n, whose details are skipped here,

and are expressed as

$$B_i(s_t|n) = \begin{cases} \frac{s_t - \kappa_i}{(\kappa_{i+n-1} - s_t)} B_i(s_t|n-1) + \frac{\kappa_{i+n} - s_t}{\kappa_{i+n} - \kappa_i} B_{i+1}(s_t|n-1), & \kappa_i \leq s_t < \kappa_{i+n} \\ 0, & \text{elsewhere,} \end{cases} \tag{3.15}$$

Efficiently computed using sparse matrix techniques, the basis functions $B_i(s_t)$ produce the design matrix that can be expressed as

$$B(s_t) = \begin{bmatrix} B_1(s_1) & B_2(s_1) & \dots & B_m(s_1) \\ \vdots & \vdots & \dots & \vdots \\ B_1(s_t) & B_2(s_t) & \dots & B_m(s_t) \end{bmatrix} \tag{3.16}$$

Example 3.4. *Figure 3.4 illustrates two scenarios of a B-spline of orders $n = 2$ in Panels (a) and (c) as well as order $n = 4$ in Panels (b) and (d), to approximate the logistic function in equation (2.13) with a variable s_t and parameters $\gamma = 0.3$ and $c = \bar{s}_t$. The graphs in Panels (a) and (b) show the $m = 4$ and $m = 6$ B-spline functions, respectively, those from the design matrix (3.16), with interior knots at positions 20, 40 and 60 of s_t, identified in the figure as dashed vertical lines. The shape of the curve changes from straight lines to cubic curves, once the order n increases from 2 to 4. Panel (c) presents the four polynomial straight lines specified in each interval [0, 20], [20, 40], [40, 60] and [60, 80] to approximate the logistic function. This B-spline function with $r = 1$ is an alternative approximation to that from the TPB spline function with $r = 1$ shown in Figure 3.3. Panel (d) shows the B-splines approximation with degree $r = 3$. The gain in increasing the degree of the polynomials is noticeable, since the approximation with $r = 3$ is closer to the*

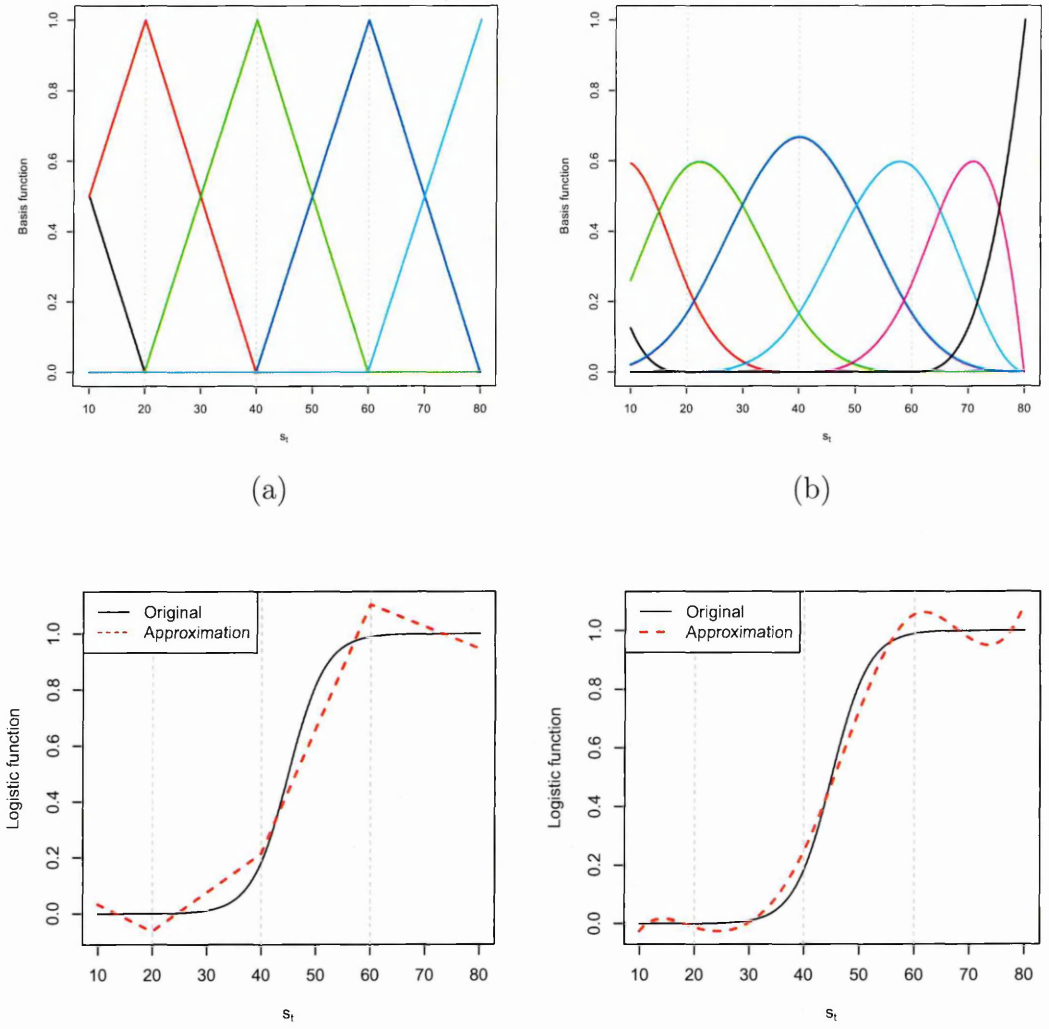


Figure 3.4: B-spline basis functions of (a) degree $r = 1$ and (b) degree $r = 3$, of a logistic function at variable $s_t, \gamma = 0.3$ and $c = \bar{s_t}$, as well as the corresponding B-spline approximations (c) and (d), respectively.

function.

3.6 Solution for overdetermined systems of polynomial equations

This section presents a method for finding a solution for the overdetermined systems of polynomial equations, obtained during proposing DBSTAR models in the next Chapter 4. Systems of polynomial equations generally hold for more problems than systems of linear equations. Overdetermined systems are those with several equations and a restricted number of unknowns, i.e., the systems present more equations than unknowns.

There are many methods for finding solutions for overdetermined systems of polynomial equations in the literature. For a review, see Ortega and Rheinboldt (1970). In this thesis, we adopt a method that is applied for minimising a function $f(.)$ subject to constraints $h(.) = b$ in an optimisation problem. This method uses the Lagrange multipliers technique.

Consider a system with q equations on n unknowns $x_1, \ldots, x_n$, where $q > n$, and which are functions $\mathbb{R}^n \to \mathbb{R}^1$, $f_{1,\ldots,q}(x_1, \ldots, x_n)$. The Lagrange function is defined as

$$\begin{aligned} \Lambda(x_1, \ldots, x_n, \lambda) &= f_{1,\ldots,q}(x_1, \ldots, x_n) + \lambda \left[h(x_1, \ldots, x_n)\right] \\ h(x_1, \ldots, x_n) &= b \end{aligned} \tag{3.17}$$

where λ is called Lagrange multiplier, $h(x_1, \ldots, x_n)$ is the function defining the

constraints on $x_1, \ldots, x_n$ and b is the value of the function with constraints.

The first derivative of the Lagrange function $\Lambda(x_1, \ldots, x_n, \lambda)$ with respect to the Lagrange multiplier λ is set to zero, denoted by $\nabla_\lambda \Lambda(x_1, \ldots, x_n, \lambda) = 0$, and needs to be solved, which implies $h(x_1, \ldots, x_n) = b$. By differentiating the constraints $h(x_1, \ldots, x_n)$ with respect to each unknown $x_i, i = 1, \ldots, n$ and giving consistent initial values to $x_1, \ldots, x_n$ and λ, a minimisation algorithm, such as descend methods or Newton's method (Ortega and Rheinboldt, 1970), is needed for which the values of the Lagrange function iteratively decreases at each stage to find the solution for each unknown x_i in the overdetermined systems of polynomial equations.

Example 3.5. *Consider a simple illustration of this method with an overdetermined systems of linear equations. An example of an overdetermined system of polynomial equations is presented in Chapter 4, in which has similar idea to this example described here with linear equations.*

A stationary linear AR model of order p from (1.2) can have its parameters estimated with Lagrange multiplier method. The model with $p = 1$ *is stated as*

$$Y_t = \alpha_0 + \alpha_1 y_{t-1} + \epsilon_t.$$

Assuming that the number of observations is greater than two ($T > 2$)*, this AR(1) model forms an overdetermined system with* T *equations and* 2 *unknowns* (α_0 *and*

α_1), as follows

$$\begin{aligned} Y_1 &= \alpha_0 + \alpha_1 y_0 + \epsilon_1 \\ Y_2 &= \alpha_0 + \alpha_1 y_1 + \epsilon_2 \\ &\vdots \\ Y_T &= \alpha_0 + \alpha_1 y_{T-1} + \epsilon_T \end{aligned}$$

To estimate the parameters α_0 and α_1, given a set of data points $y_t, t = 1, \ldots, T$, we need to find the value for each parameter $\alpha_i, i = 0, 1$, that minimises the sum of squared residuals, i.e., $\sum_{t=1}^{T}(y_t - \alpha_0 - \alpha_1 y_{t-1})^2$. Notice that this is similar to a linear least squares problem. However, suppose that the estimated values of both parameters have to be identical, i.e., $\hat{\alpha_0} = \hat{\alpha_1}$. With this constraint, the Lagrange function in (3.17) in this case is as

$$\begin{aligned} \Lambda(\alpha_0, \alpha_1, \lambda) &= \sum_{t=1}^{T}(y_t - \alpha_0 - \alpha_1 y_{t-1})^2 + \lambda\left(\alpha_0^2 - \alpha_1^2\right) \\ \alpha_0^2 - \alpha_1^2 &= 0. \end{aligned} \tag{3.18}$$

Then, a minimisation algorithm can be applied to find the solution for parameters α_i.

3.7 Summary

In this chapter, fundamental mathematical and statistical backgrounds were described in order to set the scene by defining the key modelling concepts to then

propose DBSTAR models in the next Chapters 4 and 5.

The models surveyed in the previous Chapter 2 have all the parameters fixed in time. The idea is to rewrite them into the DLM form so that the resulting models will allow parameters to change over time. Approximation methods such as the Taylor Series expansion or Splines for the transition function are required in this stage. The method used for estimating the parameters in such situation is the Kalman Filter technique. Overdetermined systems of polynomial equations are formed when approximating the transition function. Therefore, after obtaining the estimates of polynomial coefficients using the Kalman Filter algorithm, Lagrange Multiplier method is required to find estimated values of the original parameters $(\underline{\alpha}_1, \underline{\alpha}_2, \gamma, c)$.

In summary, the estimation algorithm step-by-step is:

- Take the STAR model and its transition function
- Apply an approximation method to its transition function
- Rewrite the polynomial into the DLM form
- Estimate its parameters using the Kalman Filter algorithm
- Obtain estimates for the original parameters of the STAR model using the Lagrange Multiplier method

This estimation algorithm is fully illustrated in Section 4.2.1.

Chapter 4

Dynamic Bayesian STAR models

The aim of this chapter is to introduce a class of Dynamic Bayesian STAR (DBSTAR) models which generalises classical STAR models of Chan and Tong (1986a) and CBSTAR models of Lopes and Salazar (2005). DBSTAR models allow time-dependent parameters including observational variance and are suitable for modelling local non-stationarity and heteroscedastic processes. DBSTAR models make use of the Kalman filter for fast sequential parametric estimation and are thus suitable for high frequency time series.

We start by defining DBSTAR models that can be customised for different transition functions as well as for different approaches for approximating those as we shall see. In particular, Taylor series expansions used to approximate logistic transition functions define Taylor DBSTAR models. Similarly, the use of splines bases to approximate transition functions will define Splines DBSTAR models. We also show approaches for retaining parametric interpretability related to STAR models that involve solving systems of polynomial equations.

4.1 Model definition

We start by rewriting the classical STAR models in (2.12) as follows

$$\begin{aligned} Y_t &= \underline{z}_t'\underline{\alpha}_1\pi(s_t;\gamma,c) + \underline{z}_t'\underline{\alpha}_2[1-\pi(s_t;\gamma,c)] + \epsilon_t \\ &= \underline{z}_t'\underline{\alpha}_1\pi(s_t;\gamma,c) + \underline{z}_t'\underline{\alpha}_2 - \underline{z}_t'\underline{\alpha}_2\pi(s_t;\gamma,c) + \epsilon_t \\ &= \underline{z}_t'\underline{\alpha}_2 + \underline{z}_t'(\underline{\alpha}_1-\underline{\alpha}_2)\pi(s_t;\gamma,c) + \epsilon_t \\ &= \underline{z}_t'\underline{\phi}_1 + \underline{z}_t'\underline{\phi}_2\pi(s_t;\gamma,c) + \epsilon_t \quad ; \quad \epsilon_t \sim N(0,\sigma^2) \end{aligned} \tag{4.1}$$

where $\underline{\phi}_1 = \underline{\alpha}_2$ and $\underline{\phi}_2 = (\underline{\alpha}_2 - \underline{\alpha}_1)$ are $(p+1)$-dimensional vectors with elements ϕ_{ji} $(j = 0,1,\ldots,p, i = 1,2)$ representing autoregressive coefficients associated with each component of $\underline{z}_t' = (1, y_{t-1},\ldots,y_{t-p})$. Recall that $\pi(.)$ is a nonlinear smooth transition function in the range $[0,1]$ with parameters $\gamma \in \mathbb{R}^+$ and $c \in \mathbb{R}$, respectively, and $s_t \in \mathbb{R}$ defined as a transition variable (usually in practice being either an external variable or lagged dependent variable y_{t-d}, where d is a delay parameter), and ϵ_t is independent and identically normally distributed.

For a time series process Y_t, at time $t = 1,2,\ldots,T$, STAR models of order p in (4.1) can be rewritten as,

$$Y_t = \underline{z}_t'\underline{\phi}_{1t} + \underline{z}_t'\underline{\phi}_{2t}\pi(s_t;\gamma_t,c_t) + \epsilon_t \quad ; \quad \epsilon_t \sim N(0,\sigma_t^2) \tag{4.2}$$

where $\underline{\phi}_{1t} = (\phi_{10t}, \phi_{11t},\ldots,\phi_{1pt})$ and $\underline{\phi}_{2t} = (\phi_{20t}, \phi_{21t},\ldots,\phi_{2pt})$ are $(p+1)$-dimensional vectors with elements representing time-dependent autoregressive coefficients, σ_t^2 is a time-dependent variance, and $\pi(s_t;\underline{\psi}_t)$ is a dynamic nonlinear smooth transition

function in the range $[0, 1]$ at each t given by

$$\pi(s_t; \underline{\psi}_t) = [1 + \exp\{-\gamma_t(s_t - c_t)\}]^{-1} \tag{4.3}$$

The autoregressive coefficients, the nonlinear parameters of the transition function and the variance are now allowed to change over time. In such a way, they become time-dependent parameters $\underline{\phi}_{1t}, \underline{\phi}_{2t}, \gamma_t, c_t$ and σ_t^2 likewise the transition function $\pi(s_t; \gamma_t, c_t)$. Extending STAR models by allowing both the parameters and transition function to be dynamic means that, at each time t, the transition between AR regimes may happen with different degrees of smoothness and/or at different thresholds. Further to that, the variance can also be adapted each time a new observation becomes available. This approach may better reflect reality in many applications. However, these extensions have implications in terms of estimation, since the usual linear and nonlinear least squares methods adopted by classical STAR models cannot be used for this approach. The reason is that classical STAR models are static approaches that require the processes to be stationary. MCMC algorithms adopted by CBSTAR models cannot be used either, since MCMC approach is computationally intensive for sequential estimation and convergence of the chains may be slow.

DBSTAR models avoid these estimation problems by reformulating the model with time-dependent parameters in (4.2) as DLMs and, thus, using the Kalman filter algorithm for estimation parameters and producing forecasting. DBSTAR models allow computations to be made recursively according to the observed data arriving sequentially over time, incorporating new information and summarising the past history into the dynamic parameters.

Example 4.1. *Suppose that we are interested in modelling Electricity load, Y_t, in a specific region and we have Temperature as a transition variable, s_t, as we shall see in Chapter 7. Different patterns through time are identified for both y_t and s_t. Any transition occurring during Winter time may not be the same as in Summer. Hence, smooth transitions between regimes are naturally expected to occur at different temperatures, according to the season, so the values of c in $\pi(s_t;\gamma,c)$ may change. In addition, smoother transitions may occur during Winter than in Summer or vice-versa, therefore the values of γ may also change over time.*

DBSTAR models use approximation methods, such as polynomial approximations or splines, during the model formulation stage to represent a generic dynamic transition function $\pi(s_t;\gamma_t,c_t)$ whilst both the classical and Bayesian STAR models reviewed in Chapter 2 use approximation methods, such as nonlinear least squares or MCMC, during the estimation stage.

DBSTAR models are thus defined based upon the use of approximation methods to a generic dynamic transition function $\pi(s_t;\gamma_t,c_t)$ shown in equation (4.2), as follows. A general form of a STAR model using an approximation $\tilde{\pi}(s_t;\gamma_t,c_t)$ to a transition function $\pi(s_t;\gamma_t,c_t)$ can be expressed as

$$\begin{aligned} Y_t &= \underline{z}_t'\underline{\phi}_{1t} + \underline{z}_t'\underline{\phi}_{2t}\pi(s_t;\gamma_t,c_t) + \epsilon_t \\ &\cong \underline{z}_t'\underline{\phi}_{1t} + \underline{z}_t'\underline{\phi}_{2t}\tilde{\pi}(s_t;\gamma_t,c_t) + \xi_t \end{aligned} \tag{4.4}$$

where $\xi_t = \epsilon_t + \underline{z}_t'\underline{\phi}_{2t}R_t(s_t;\gamma_t,c_t)$, with $R_t(s_t;\gamma_t,c_t) = \pi(s_t;\gamma_t,c_t) - \tilde{\pi}(s_t;\gamma_t,c_t)$ the remainder term from the polynomial approximations or splines. The approximation

is expected to be as close to the dynamic transition function as possible so that $\lim_{r\to\infty} R_t(s_t;\gamma_t,c_t) = 0$, therefore for good approximations, this remainder term does not affect the properties of the errors ξ_t and hence they should have approximately zero mean and variance σ_t^2, $\xi_t \sim N(0,\sigma_t^2)$.

The polynomial approximations will help to characterise $\tilde{\pi}(s_t;\gamma_t,c_t)$, so that particular cases of approximate STAR models can be specified before defining DBSTAR models. Notice that we focus on some of those approximation methods presented in Section 3.4, but other methods could also be used.

The use of polynomial approximations, such as Taylor series expansions, for the dynamic transition function $\pi(s_t;\gamma_t,c_t)$ are reasonably the simplest approaches for approximating nonlinear functions. Generally, a smooth transition function $\pi(s_t;\gamma_t,c_t)$ is specified first, e.g., an S-shape curve in relation to nonlinearity of the underlying process. The first-order logistic function defined in equation (2.13) is the most commonly used function. Correspondingly, we find its Taylor approximation $\tilde{\pi}(s_t;\gamma_t,c_t)$, as shown in equation (3.10). Notice that, despite focusing on the logistic function, the results here are valid for any other transition function as mentioned in Chapter 2.

Consider a generic polynomial approximation $P_r(s_t) = \tilde{\pi}(s_t;\gamma_t,c_t)$ for the transition function $\pi(s_t;\gamma_t,c_t)$ as in equation (3.7) given by

$$P_r(s_t) = \sum_{i=0}^{r} \alpha_i B_i(s_t)$$

where $\alpha_i, i = 0,\ldots,r$, are coefficients of the polynomial which determine the shape of the polynomial approximation. The general form of the model introduced in (4.4)

becomes

$$Y_t \cong \underline{z}_t'\underline{\phi}_{1t} + \underline{z}_t'\underline{\phi}_{2t}\left[\sum_{i=0}^{r}\alpha_{it}B_i(s_t)\right] + \xi_t \quad ; \quad \xi_t \sim N(0,\tilde{\sigma}_t^2) \tag{4.5}$$

where the polynomial degree r and the form of $B_i(.), i = 0,\ldots,r$, with $B_0(s_t) = 1$, need to be specified, as we shall see. Notice that $\lim_{r\to\infty}[\pi(s_t;\gamma_t,c_t) - \tilde{\pi}(s_t;\gamma_t,c_t)] = 0$ should be valid for a good approximation method subject to $\xi_t \to \epsilon_t$, therefore $E[\xi_t] = E[\epsilon_t] = 0$ and $\tilde{\sigma}_t^2 = Var[\xi_t] = Var[\epsilon_t] = \sigma_t^2$.

Example 4.2. *For a polynomial degree $r = 3$ and $B_i(s_t) = s_t^i, i = 1,2,3$, set as power functions, as defined in equation (3.7), the model defined in (4.5) becomes*

$$\begin{aligned}
Y_t &\cong \underline{z}_t'\underline{\phi}_{1t} + \underline{z}_t'\underline{\phi}_{2t}\left[\sum_{i=0}^{3}\alpha_{it}s_t^i\right] + \xi_t \\
&\cong \underline{z}_t'\underline{\phi}_{1t} + \underline{z}_t'\underline{\phi}_{2t}\left[\alpha_{0t} + \alpha_{1t}s_t + \alpha_{2t}s_t^2 + \alpha_{3t}s_t^3\right] + \xi_t \\
&\cong \underline{z}_t'\underline{\phi}_{1t} + \alpha_{0t}\underline{z}_t'\underline{\phi}_{2t} + \alpha_{1t}\underline{z}_t'\underline{\phi}_{2t}s_t + \alpha_{2t}\underline{z}_t'\underline{\phi}_{2t}s_t^2 + \alpha_{3t}\underline{z}_t'\underline{\phi}_{2t}s_t^3] + \xi_t \\
&\cong \underline{z}_t'\underline{\theta}_{1t} + (\underline{z}_t's_t)\underline{\theta}_{2t} + (\underline{z}_t's_t^2)\underline{\theta}_{3t} + (\underline{z}_t's_t^3)\underline{\theta}_{4t} + \xi_t
\end{aligned} \tag{4.6}$$

where $\underline{\theta}_{1t} = \underline{\phi}_{1t} + \alpha_0\underline{\phi}_{2t}, \underline{\theta}_{2t} = \alpha_1\underline{\phi}_{2t}, \underline{\theta}_{3t} = \alpha_2\underline{\phi}_{2t}$ and $\underline{\theta}_{4t} = \alpha_3\underline{\phi}_{2t}$ are the coefficients associated with both the autoregressive part of the model and the transition variable s_t. Notice that $\alpha_{it}, i = 0,\ldots,3$ are scalars and $\underline{\phi}_{1t}$ and $\underline{\phi}_{2t}$ are vectors so the product of them hold.

The following Section 4.2 defines Taylor DBSTAR models.

4.2 Taylor DBSTAR Models

To define Taylor DBSTAR models, with a Taylor series expansion of degree r, we first need to approximate the dynamic nonlinear transition function with the most appropriate Taylor series expansion. The degree r is assumed *fixed.* In cases where data are initially available, r can be determined, for example, by a grid search approach for optimal values that minimise a forecasting error statistic. Alternatively, model selection criteria such as those described in Section 3.3, e.g., the Bayes' factor method proposed by West (1986) for monitoring a model's predictive performance via the likelihood function could also be used.

Adopting the logistic function as the dynamic transition function $\pi_t(.)$ and its Taylor series approximation, the model in (4.5) becomes

$$Y_t = \underline{z}_t'\underline{\theta}_{1t} + (\underline{z}_t's_t)\underline{\theta}_{2t} + (\underline{z}_t's_t^2)\underline{\theta}_{3t} + \ldots + (\underline{z}_t's_t^r)\underline{\theta}_{(r+1)t} + \xi_t \tag{4.7}$$

where for $i = 1, \ldots, r+1, \underline{\theta}_{it} = \beta_{it}(\gamma_t, c_t)\underline{\phi}_t$, and $\beta_{it}(\gamma_t, c_t)$ are polynomial functions of γ_t and c_t obtained from the i-th coefficient of the Taylor series expansion, that is

$$\beta_{it} = \left.\frac{\partial^i \pi(s_t; \gamma_t, c_t)}{\partial s_t^i \partial \gamma_t^i \partial c_t^i}\right|_{s_t=s_0, \gamma_t=\gamma_0, c_t=c_0} \tag{4.8}$$

where, for $i = 0$, $\beta_{0t} = \pi(s_0; \gamma_0, c_0)$, with s_0, γ_0 and c_0 being constant values around which the Taylor series expansion is to be carried out. In addition, $E[\xi_t] = 0$ and $Var[\xi_t] = \sigma_t^2$.

Notice that the Taylor series expansion approximates $\pi(.)$ better in the vicinities of s_0, γ_0 and c_0, so that, at each time t, those values can be conveniently chosen

to improve the approximation (e.g. they can be chosen to be closer to the turning points of the transition function when it is known that a regime change is about to happen).

Example 4.3. *For a Taylor series of degree $r = 3$ expanded around the point $s_t = c_t$, the model defined in (4.7) becomes*

$$y_t \cong \underline{z}_t'\underline{\theta}_{1t} + (\underline{z}_t's_t)\underline{\theta}_{2t} + (\underline{z}_t's_t^2)\underline{\theta}_{3t} + (\underline{z}_t's_t^3)\underline{\theta}_{4t} + \xi_t \tag{4.9}$$

where the Taylor series expansion's coefficients as in (4.8) are

$$\beta_{1t} = \frac{24 - 12\gamma_t c_t + \gamma_t^3 c_t^3}{48} \tag{4.10}$$

$$\beta_{2t} = \frac{12\gamma_t - 3\gamma_t^3 c_t^2}{48} \tag{4.11}$$

$$\beta_{3t} = \frac{3\gamma_t^3 c_t}{48} \tag{4.12}$$

$$\beta_{4t} = -\frac{\gamma_t^3}{48} \tag{4.13}$$

and, consequently,

$$\underline{\theta}_{1t} = \frac{24\underline{\phi}_{2t} - 12\gamma_t c_t \underline{\phi}_{2t} + \gamma_t^3 c_t^3 \underline{\phi}_{2t} + 48\underline{\phi}_{1t}}{48} \tag{4.14}$$

$$\underline{\theta}_{2t} = \frac{12\gamma_t \underline{\phi}_{2t} - 3\gamma_t^3 c_t^2 \underline{\phi}_{2t}}{48} \tag{4.15}$$

$$\underline{\theta}_{3t} = \frac{3\gamma_t^3 c_t \underline{\phi}_{2t}}{48} \tag{4.16}$$

$$\underline{\theta}_{4t} = -\frac{\gamma_t^3 \underline{\phi}_{2t}}{48} \tag{4.17}$$

where $\underline{\theta}_{it}, i = 1, \ldots, 4$, are the coefficients associated with both the autoregressive part of the model and the transition variable s_t.

For each time t, Taylor DBSTAR(r,p) models with p as the autoregressive order, r as the Taylor series degree and the quadruple $\Delta_t = \{\underline{F}_t, \mathbf{G}_t, V_t, \mathbf{W}_t\}$, where $\underline{F}_t$ is a known $(r+1)(p+1)$-dimensional vector, $\mathbf{G}_t$ is a known $(r+1)(p+1)\times(r+1)(p+1)$ matrix, V_t is an unknown scalar and $\mathbf{W}_t$ is an unknown $(r+1)(p+1)\times(r+1)(p+1)$ matrix, are uniquely defined by the following three probabilities:

the conditional distribution of the observation,

$$(Y_t \mid \underline{\theta}_t) \sim N\left(\underline{F}_t'\underline{\theta}_t, V_t\right) \tag{4.18}$$

the conditional distribution of the parameters (state),

$$\left(\underline{\theta}_t \mid \underline{\theta}_{t-1}\right) \sim N\left(\mathbf{G}_t\underline{\theta}_{t-1}, \mathbf{W}_t\right). \tag{4.19}$$

and the initial information of the parameters,

$$(\underline{\theta}_0 \mid D_0) \sim N\left(\underline{m}_0, \mathbf{C}_0\right), \tag{4.20}$$

Using the components of Δ_t as well as a dynamic transition nonlinear function $\pi(s_t;\gamma_t,c_t)$ with real values in the interval $[0,1]$, where s_t is a transition variable, γ_t is a dynamic parameters associated with smoothness and c_t is the threshold parameter, Taylor DBSTAR(r,p) models are specified by rewriting the model in the equation (4.7) into the observational process Y_t conditional on the parameter vector $\underline{\theta}_t$ in (4.18), where $\underline{F}_t' = [\underline{z}_t', \underline{z}_t's_t, \ldots, \underline{z}_t's_t^r]$ with a known $(r+1)(p+1)$-dimensional vector of polynomial regression variables $s_t^i\underline{z}_t'$ $(i = 0,1,\ldots,r)$ with $\underline{z}_t' = (1, y_{t-1}, \ldots, y_{t-p})$;

$\underline{\theta}_t$ is the state vector containing parameters associated with the components of $\underline{F}_t$, i.e, $\underline{\theta}_t' = (\underline{\theta}_{1t}, \ldots, \underline{\theta}_{(r+1)t})$.

The $(r+1)(p+1)$-dimensional state vector $\underline{\theta}_t$ in (4.19) has a conditional multivariate Student-t distribution with n_{t-1} degrees of freedom, with $\mathbf{G}_t$ as the state evolution matrix with elements g_{ijt} (for row i and column j) chosen according to the desired structural form of association between $\underline{\theta}_t$ and $\underline{\theta}_{t-1}$. In the case that no structural form is known, the random walk is used by setting $\mathbf{G}_t = \mathbf{I}$, the identity matrix. Other structural forms of $\mathbf{G}_t$ are given in Chapter 5 for modelling periodic processes, for example. Furthermore, $\mathbf{W}_t$ is the state covariance matrix, for which a discount factor δ_W, satisfying the condition $0 < \delta_W \leq 1$, is considered, as follows,

$$\mathbf{W}_t = \left(\frac{1-\delta_W}{\delta_W}\right) \mathbf{G}_t \mathbf{C}_{t-1} \mathbf{G}_t'.$$

where $\mathbf{C}_{t-1}$ is a prior covariance matrix for $\underline{\theta}_t$. Notice that in the case when $\delta_W = 1$, the parameters have zero variance ($\mathbf{W}_t = \mathbf{0}$).

The observational variance V_t is defined by an appropriately chosen variance discounting technique. V_t is assumed to change but only slowly and steadily over time to avoid potential unpredictable behaviour that can lead to loss of analytical tractability. In this way, the algorithm for obtaining the prior-to-posterior updating is adapted to this case of heteroscedasticity.

Example 4.4. *At each time t, a simple Taylor DBSTAR(3,1) model (with no intercept, Taylor series degree $r = 3$ and autoregressive order $p = 1$) and transition variable $s_t = y_{t-1}$ is specified by the distribution of the observational process in equation (4.18) and the distribution of the state vector in equation (4.19), with the com-*

ponents of the quadruple Δ_t *as a 4-dimensional vector* $\underline{F}_t' = [y_{t-1}, y_{t-1}^2, y_{t-1}^3, y_{t-1}^4]$, *a* 4×4 *evolution matrix* $\mathbf{G}_t = \mathbf{I}$ *as the identity matrix and a* 4×4 *state covariance matrix* $\mathbf{W}_t$ *specified by a value of* δ_W. *Also,* $\underline{\theta}_t = [\theta_{0t}, \theta_{1t}, \theta_{2t}, \theta_{3t}]$ *is a 4-dimensional state vector containing parameters associated with each element of* $\underline{F}_t$ *and they are functions equal to those given in Example 4.3. Notice that in the case that* θ_{1t} , θ_{2t} *and* θ_{3t} *are zero, the Taylor DBSTAR(3,1) model is reduced to a linear AR(1) model written in DLM form.*

In the following Section 4.2.1, we show that STAR models are particular cases of DBSTAR models. We propose methods for obtaining the corresponding parameters of STAR models to enable us to interpret them accordingly. One advantage of DBSTAR models over both the classical and Bayesian STAR models is that we are able to obtain estimates of the corresponding parameters and interpret them at each time t, whilst STAR models have static parameters for the whole period.

4.2.1 Estimating AR and transition parameters

As we have seen in Section 4.2, the state parameters $\underline{\theta}_{it}, i = 1, 2, \ldots, (r+1)(p+1)$, of Taylor DBSTAR$(r, p)$ models are compounded of coefficients that are polynomial functions of the corresponding parameters in STAR models, i.e, the AR coefficients $\underline{\phi}_{1t}$ and $\underline{\phi}_{2t}$, the smoothness and threshold parameters, γ_t and c_t, respectively. Additionally, at each time t, after observing $Y_t = y_t$, the first two moments (but not the probability distribution) of the parameters $\underline{\phi}_{1t}, \underline{\phi}_{2t}, \gamma_t$ and c_t can be obtained from the first two moments of the posterior Student-t distribution of $\underline{\theta}_t$ by solving a system of $(r+1)(p+1)$ polynomial equations on $(2p+4)$ variables for the posterior mean

$E(\underline{\theta}_t|D_t) = \underline{m}_t$ and another system for the posterior variance $Var(\underline{\theta}_t|D_t) = \text{diag}(\mathbf{C}_t)$. A description of a method available to solve systems of polynomial equations (SPE for short) for the first two moments of Taylor DBSTAR models is given in detail in Section 3.6.

The SPEs are originally obtained from the model in equation (4.7). Each of the $(r+1)(p+1)$ coefficients θ_{it} is a polynomial function of degree r of their $(2p+4)$ arguments $\underline{\phi}_{1t}, \underline{\phi}_{2t}, \gamma_t$ and c_t only. In fact, once we have obtained the filtering distribution moments through the Kalman filter, we will be solving two SPEs at each time t, namely (i) for the posterior means

$$\begin{aligned} m_{1t} &= f_{1t}(\underline{\phi}_{1t}, \underline{\phi}_{2t}, \gamma_t, c_t) \\ m_{2t} &= f_{2t}(\underline{\phi}_{1t}, \underline{\phi}_{2t}, \gamma_t, c_t) \\ &\vdots \\ m_{(r+1)(p+1)t} &= f_{(r+1)(p+1)t}(\underline{\phi}_{1t}, \underline{\phi}_{2t}, \gamma_t, c_t) \end{aligned} \tag{4.21}$$

and (ii) for the posterior variance

$$\begin{aligned} \mathbf{C}_{1t} &= f^*_{1t}(\underline{\phi}_{1t}, \underline{\phi}_{2t}, \gamma_t, c_t) \\ \mathbf{C}_{2t} &= f^*_{2t}(\underline{\phi}_{1t}, \underline{\phi}_{2t}, \gamma_t, c_t) \\ &\vdots \\ \mathbf{C}_{(r+1)(p+1)t} &= f^*_{(r+1)(p+1)t}(\underline{\phi}_{1t}, \underline{\phi}_{2t}, \gamma_t, c_t) \end{aligned} \tag{4.22}$$

Example 4.5. *The model in Example 4.4 has the following SPEs for the posterior*

means, at each time t, corresponding to (i) as defined in (4.21)

$$
\begin{aligned}
m_{1t} &= \frac{24\hat{\phi}_{2t} - 12\hat{\gamma}_t\hat{c}_t\hat{\phi}_{2t} + \hat{\gamma}_t^3\hat{c}_t^3\hat{\phi}_{2t} + 48\hat{\phi}_{1t}}{48} \\
m_{2t} &= \frac{12\hat{\gamma}_t\hat{\phi}_{2t} - 3\hat{\gamma}_t^3\hat{c}_t^2\hat{\phi}_{2t}}{48} \\
m_{3t} &= \frac{3\hat{\gamma}_t^3\hat{c}_t\hat{\phi}_{2t}}{48} \\
m_{4t} &= -\frac{\hat{\gamma}_t^3\hat{\phi}_{2t}}{48}
\end{aligned}
$$

where the estimated posterior means $m_{it}, i = 1, \ldots, 4$ are obtained from the Kalman filtering distributions, and SPEs for the posterior variances corresponding to (ii) as defined in (4.22)

$$
\begin{aligned}
C_{1t} &= \frac{24Var(\phi_{2t}) - 12Var(\gamma_t)Var(c_t)Var(\phi_{2t}) + Var(\gamma_t)^3Var(c_t)^3Var(\phi_{2t}) + 48Var(\phi_{1t})}{48} \\
C_{2t} &= \frac{12Var(\gamma_t)Var(\phi_{2t}) - 3Var(\gamma_t)^3Var(c_t)^2Var(\phi_{2t})}{48} \\
C_{3t} &= \frac{3Var(\gamma_t)^3Var(c_t)Var(\phi_{2t})}{48} \\
C_{4t} &= -\frac{Var(\gamma_t)^3Var(\phi_{2t})}{48}
\end{aligned}
$$

where the estimated posterior variances $C_{it}, i = 1, \ldots, 4$ are also obtained from the Kalman filtering distributions.

It is required to use mathematical methods for finding solutions for these SPEs. The Lagrange Multipliers (LM) method, described in Section 3.6, finds solutions for unknowns subject to constraints. In order to associate the SPEs (4.21) and (4.22) with optimisation problems, so the LM method can be applied, first, disturbance terms $\eta_{1t}, \ldots, \eta_{(r+1)(p+1)t}$ are added individually to each $(r+1)(p+1)$ polynomial

equation, at each time t, as follows

$$\begin{aligned} m_{1t} &= f_{1t}(\underline{\phi}_{1t}, \underline{\phi}_{2t}, \gamma_t, c_t) + \eta_{1t} \\ m_{2t} &= f_{2t}(\underline{\phi}_{1t}, \underline{\phi}_{2t}, \gamma_t, c_t) + \eta_{2t} \\ &\vdots \\ m_{(r+1)(p+1)t} &= f_{(r+1)(p+1)t}(\underline{\phi}_{1t}, \underline{\phi}_{2t}, \gamma_t, c_t) + \eta_{(r+1)(p+1)t} \end{aligned} \tag{4.23}$$

and (ii)

$$\begin{aligned} \mathbf{C}_{1t} &= f^*_{1t}(\underline{\phi}_{1t}, \underline{\phi}_{2t}, \gamma_t, c_t) + \eta^*_{1t} \\ \mathbf{C}_{2t} &= f^*_{2t}(\underline{\phi}_{1t}, \underline{\phi}_{2t}, \gamma_t, c_t) + \eta^*_{2t} \\ &\vdots \\ \mathbf{C}_{(r+1)(p+1)t} &= f^*_{(r+1)(p+1)t}(\underline{\phi}_{1t}, \underline{\phi}_{2t}, \gamma_t, c_t) + \eta^*_{(r+1)(p+1)t}, \end{aligned} \tag{4.24}$$

Notice that both SPEs (4.23) and (4.24) have same aspects as the system of linear equations shown in Example 3.5. Therefore, the LM method finds the solution that minimises $\sum_{i=1}^{(r+1)(p+1)} \eta_{it}^2$ for the system (i), i.e., we obtain the posterior means of each corresponding parameter given by the real part of a polynomial function of degree r of their $(r+1)(p+1)$ location point estimates $g_{jt}(\underline{m}_t), j = 1, \ldots, (2p+4)$, i.e., $\hat{\underline{\phi}}_{1t} = \text{Re}\{g_{1t}(\underline{m}_t)\}$, $\hat{\underline{\phi}}_{2t} = \text{Re}\{g_{2t}(\underline{m}_t)\}$, $\hat{\gamma}_t = \text{Re}\{g_{3t}(\underline{m}_t)\}$ and $\hat{c}_t = \text{Re}\{g_{4t}(\underline{m}_t)\}$, where $\hat{\psi}_t = E(\psi_t|D_t)$ and Re{} is the real part of the number; and similarly for the system (ii), we obtain the posterior variances of each original parameter given by the real part of a polynomial function $g^*_{jt}(\mathbf{C}_{it})$, i.e., $Var(\underline{\phi}_{1t}|D_t) = \text{Re}\{g^*_{1t}(\mathbf{C}_{it})\}$,

$Var(\underline{\phi}_{2t}|D_t) = \text{Re}\{g^*_{2t}(\mathbf{C}_{it})\}$, $Var(\gamma_t|D_t) = \text{Re}\{g^*_{3t}(\mathbf{C}_{it})\}$ and $Var(c_t|D_t) = \text{Re}\{g^*_{4t}(\mathbf{C}_{it})\}$.

Example 4.6. *Suppose that, at time $t = 2$, the Kalman filter algorithm assessed posterior means $m_{12} = 0.20, m_{22} = 0.34, m_{32} = 0.17$ and $m_{42} = -0.06$. By using the LM method, we can find the values of $\hat{\phi}_{12}, \hat{\phi}_{22}, \hat{\gamma}_2$ and $\hat{c}_2$, i.e. posterior means for each parameter $\phi_{1t}, \phi_{2t}, \gamma_t$ and c_t, at time $t = 2$.*

In this example, it took half a second to run the LM method using the software Mathematica version 9, that gave $\hat{\phi}_{12} = 0.28, \hat{\phi}_{22} = -1.87, \hat{\gamma}_2 = 1.18$ and $\hat{c}_2 = 0.92$. Those were the values that minimised the Lagrange function, i.e., the sum of squared disturbance with the constraint that one of the parameters is positive $\lambda_2 > 0$, thus, $\sum_{i=1}^{4} \eta_{i2}^2 = 2 \times 10^{-29}$. This can be considered a very small number and therefore can be considered negligible. Consequently, each η_{i2} is also considered very small disturbance, which means that the estimated values $\hat{\phi}_{12}, \hat{\phi}_{22}, \hat{\gamma}_2$ and $\hat{c}_2$ satisfy the conditions of the SPEs.

In summary, the estimation algorithm step-by-step is:

- Take the STAR model from equation (4.2)

$$Y_t = \underline{z}_t'\underline{\phi}_{1t} + \underline{z}_t'\underline{\phi}_{2t}\pi(s_t; \gamma_t, c_t) + \epsilon_t \quad ; \quad \epsilon_t \sim N(0, \sigma_t^2)$$

- Take also its transition function, for example, the logistic function

$$\pi(s_t; \underline{\psi}_t) = [1 + \exp\{-\gamma_t(s_t - c_t)\}]^{-1}$$

- Apply an approximation method to its transition function and obtain the

polynomial model

$$y_t \cong \underline{z}_t'\underline{\theta}_{1t} + (\underline{z}_t's_t)\underline{\theta}_{2t} + (\underline{z}_t's_t^2)\underline{\theta}_{3t} + (\underline{z}_t's_t^3)\underline{\theta}_{4t} + \xi_t$$

- Rewrite the resulting polynomial model into the DLM form

  $\underline{F}_t' = [\underline{z}_t', \underline{z}_t's_t, \ldots, \underline{z}_t's_t^r]$, $\mathbf{G}_t = \mathbf{I}$, V_t, W_t, $\underline{\theta}_t' = (\underline{\theta}_{1t}, \ldots, \underline{\theta}_{(r+1)t})$.

- Estimate its parameters using the Kalman Filter algorithm (obtain the moments $\underline{m}_{it}$ and $\mathbf{C}_{it}, i = 1, 2, \ldots, (r+1)(p+1)$)

- Obtain estimates for the means of the original parameters $(\underline{\phi}_{1t}, \underline{\phi}_{2t}, \gamma_t, c_t)$ of the model (4.2) using the Lagrange Multiplier method to solve the following overdetermined system of polynomial equations

$$\begin{aligned}
\underline{m}_{1t} &= \frac{24\underline{\phi}_{2t} - 12\gamma_t c_t \underline{\phi}_{2t} + \gamma_t^3 c_t^3 \underline{\phi}_{2t} + 48\underline{\phi}_{1t}}{48} + \eta_{1t} \\
\underline{m}_{2t} &= \frac{12\gamma_t \underline{\phi}_{2t} - 3\gamma_t^3 c_t^2 \underline{\phi}_{2t}}{48} + \eta_{2t} \\
\underline{m}_{3t} &= \frac{3\gamma_t^3 c_t \underline{\phi}_{2t}}{48} + \eta_{3t} \\
\underline{m}_{4t} &= -\frac{\gamma_t^3 \underline{\phi}_{2t}}{48} + \eta_{4t}
\end{aligned}$$

- Obtain estimates for the variance of the original parameters $(\underline{\phi}_{1t}, \underline{\phi}_{2t}, \gamma_t, c_t)$ of the model (4.2) using the Lagrange Multiplier method to solve the following overdetermined system of polynomial equations, similarly to the previous item

$$
\begin{aligned}
\mathbf{C}_{1t} &= \frac{24\underline{\phi}_{2t} - 12\gamma_t c_t \underline{\phi}_{2t} + \gamma_t^3 c_t^3 \underline{\phi}_{2t} + 48\underline{\phi}_{1t}}{48} + \eta_{1t}^* \\
\mathbf{C}_{2t} &= \frac{12\gamma_t \underline{\phi}_{2t} - 3\gamma_t^3 c_t^2 \underline{\phi}_{2t}}{48} + \eta_{2t}^* \\
\mathbf{C}_{3t} &= \frac{3\gamma_t^3 c_t \underline{\phi}_{2t}}{48} + \eta_{3t}^* \\
\mathbf{C}_{4t} &= -\frac{\gamma_t^3 \underline{\phi}_{2t}}{48} + \eta_{4t}^*
\end{aligned}
$$

4.3 Splines DBSTAR models

The use of polynomial approximations for the dynamic nonlinear transition functions, for example, Taylor series approximations, to define Taylor DBSTAR models in the previous Section 4.2 presents some limitations in their formulations, as mentioned in Section 3.4, for example, the transition functions have to be infinitely differentiable at the real points where the approximations are performed. Although polynomial approximations are attractive techniques, these limitations may affect the model performance as the approximations are carried out in the early stages of the formulations. They could compromise either parameter estimation or forecasting performance. However, alternative ways of approximating the transition functions in DBSTAR models, which avoid limitations of Taylor series in general, are presented in this section. Their use constitutes Splines DBSTAR models.

To define Splines DBSTAR(r, k, p) models, with fixed polynomial degree r, k knots and autoregressive order p, we do not necessarily need to know beforehand what the transition function is. The use of splines will assist in determining the

underlying functional forms of the dynamic nonlinear transition functions by using pieces of polynomials. Likewise with the Taylor series degree, the spline degree r can be determined, in cases where data are initially available, for example, by a grid search approach for optimal values that minimise a forecasting error statistic. The most common choice for polynomial degree is $r = 3$, which characterises cubic splines. Alternatively, model selection criteria such as those described in Section 3.3 could also be used.

The approximation $\tilde{\pi}(s_t; \gamma_t, c_t)$ to the transition function $\pi(s_t; \gamma_t, c_t)$ in the model (4.4) can also be specified by the use of piecewise polynomials. Depending on the splines function, e.g., truncated power basis, B-Splines and others, Splines DBSTAR models can be defined accordingly. The parameters in Splines DBSTAR models are interpreted in intervals, specified by the knots which are associated with each piecewise polynomial.

Example 4.7. *For a Splines DBSTAR model using temperature as a transition variable* ($s_t = temperature_t$) *with* 1 *inner knot, say knot* $= 25°C$. *Consider then* $min(s_t) = 10°C$ *and* $max(s_t) = 35°C$, *therefore, we have then* 2 *extra knots to characterise the piecewise polynomials,* $knot_1 = 10°C$ *and* $knot_2 = 35°C$. *So, one polynomial is given for the interval* $[10°C, 25°C]$ *and another for* $[25°C, 35°C]$, *and their autoregressive coefficients are interpreted according to the interval that* s_t *lies in.*

The next subsections define Truncated Power Basis DBSTAR models and B-Splines DBSTAR models.

4.3.1 Truncated Power Basis DBSTAR models

The quadruple Δ_t, defined in the previous Section 4.2, with p as the autoregressive order, in conjunction with r as the polynomial degree, can be directly extended to accommodate k as the number of knots, so that known $(r+k+2)(p+1)$-dimensional vector $\underline{F}_t$, a known $(r+k+2)(p+1) \times (r+k+2)(p+1)$ matrix $\mathbf{G}_t$, an unknown scalar V_t and an unknown $(r+k+2)(p+1) \times (r+k+2)(p+1)$ matrix $\mathbf{W}_t$, to propose distinctively Truncated Power Basis DBSTAR models. The model in (4.5) with a suitable truncated power basis approximation as in equation (3.12) is defined as follows,

$$\begin{aligned} y_t &= \underline{\phi}_{1t}\underline{z}_t^{'} + \underline{\phi}_{2t}\underline{z}_t^{'}\left[\sum_{a=0}^{r}\beta_{at}s_t^a + \sum_{b=1}^{k}\beta_{rbt}(s_t-\kappa_b)_+^r\right] + \xi_t \\ &= \underline{\theta}_{10t}\underline{z}_t^{'} + \sum_{a=0}^{r}\underline{\theta}_{2at}\underline{z}_t^{'}s_t^a + \sum_{b=1}^{k}\underline{\theta}_{3rbt}\underline{z}_t^{'}(s_t-\kappa_b)_+^r + \xi_t \end{aligned} \tag{4.25}$$

where $\underline{\theta}_{10t} = \underline{\phi}_{1t}, \underline{\theta}_{2at} = \underline{\beta}_{at}\underline{\phi}_{2t}$ and $\underline{\theta}_{3rbt} = \underline{\beta}_{rbt}\underline{\phi}_{2t}$, with $E[\xi_t] = 0$ and $Var[\xi_t] = \sigma_t^2$. Recall that the + sign in $(s_t-\kappa_b)_+^r$ symbolises the function takes values

$$(s_t-\kappa_b)_+^r = \begin{cases} (s_t-\kappa_b)^r, & \text{for } s_t \geq \kappa_b \\ 0, & \text{for } s_t < \kappa_b. \end{cases}$$

Notice that if all elements of $\underline{\theta}_{3rbt}$ are zero, the model is analogous to a Taylor DBSTAR model, which indicates that Taylor DBSTAR models are particular cases of Truncated Power Basis DBSTAR models. In addition, if all elements of $\underline{\theta}_{2at}$ and $\underline{\theta}_{3rbt}$ are zero, the model is then reduced to a linear time-dependent AR(p) model written in DLM form.

Truncated Power Basis DBSTAR(r, k, p) models are specified by rewriting the model in the equation (4.25) into the observational process Y_t conditional on the parameter vector $\underline{\theta}_t$ via the sequentially normal distribution as in equation (4.18), where $\underline{F}_t^{'} = [\underline{z}_t^{'}, \underline{z}_t^{'}s_t, \ldots, \underline{z}_t^{'}s_t^r, z_t(s_t - \kappa_1)_+^r, \ldots, z_t(s_t - \kappa_k)_+^r]$ with a known $(r + k + 2)(p + 1)$-dimensional vector of polynomial regression variables, and $\underline{\theta}_t$ is the state vector containing parameters associated with the components of $\underline{F}_t$. The observational variance V_t is defined by an appropriately chosen variance discount technique, as shown in Section 4.2.

The $(r+k+2)(p+1)$-dimensional state vector $\underline{\theta}_t$ has a conditional multivariate Student-t distribution with n_{t-1} degrees of freedom as in equation (4.19).

Example 4.8. *For each time t, consider a simple Truncated Power Basis DBSTAR$(3, 1, 1)$ model with no intercept and transition variable $s_t = y_{t-1}$. The components are hence established as a 6-dimensional vector $\underline{F}_t^{'} = [y_{t-1}, y_{t-1}, y_{t-1}^2, y_{t-1}^3, y_{t-1}^4, y_{t-1}(y_{t-1} - \kappa_1)_+^3]$, a 6×6 evolution matrix $\mathbf{G}_t = \mathbf{I}$, the identity matrix, and a 6×6 state covariance matrix $\mathbf{W}_t$. Also, $\underline{\theta}_t = [\theta_{10t}, \theta_{20t}, \theta_{21t}, \theta_{22t}, \theta_{23t}, \theta_{331t}]$ is a 6-dimensional state vector containing parameters associated with each element of $\underline{F}_t$.*

4.3.2 B-splines DBSTAR models

The adoption of truncated power basis functions as approximations for the dynamic nonlinear transition functions to define the models in the previous Section 4.3.1 makes DBSTAR models appealing due to their simplicity to set up. Nonetheless, truncated power basis functions can suffer from numerical instability, in some cases, such there are a large number of knots. Other limitations are mentioned in Section

3.5.1. These limitations can be avoided by using sufficiently more stable similar basis functions known as B-splines.

As well as in the previous Section 4.3.1, where p is the autoregressive order and r is the splines degree, with the exception of designating k as the number of inner knots, as stated in Section 3.5.2, and $m = r+1+k$, the components in the quadruple Δ_t, defined in the previous Section 4.2, are adjusted to a known $(m+1)(p+1)$-dimensional vector $\underline{F}_t$, a known $(m+1)(p+1) \times (m+1)(p+1)$ matrix $\mathbf{G}_t$, an unknown scalar V_t and an unknown $(m+1)(p+1) \times (m+1)(p+1)$ matrix $\mathbf{W}_t$, appropriately to describe B-splines DBSTAR models.

The model (4.5) with a suitable B-splines approximation as in equation (3.12) is defined as follows,

$$\begin{aligned} y_t &= \underline{\phi}_{1t} z_t + \underline{\phi}_{2t} z_t \left[\sum_{i=1}^{m} \alpha_{it} B_i(s_t) \right] + \xi_t \\ &= \underline{\theta}_{10t} z_t + \sum_{i=1}^{m} \underline{\theta}_{2it} z_t B_i(s_t) + \xi_t \end{aligned} \tag{4.26}$$

where $\underline{\theta}_{10t} = \underline{\phi}_{1t}, \underline{\theta}_{2it} = \alpha_{it}\underline{\phi}_{2t}$, $i = 1, \ldots, m$ and $E[\xi_t] = 0$ and $Var[\xi_t] = \sigma_t^2$. If all elements of $\underline{\theta}_{2it}$ are zero, the model is then reduced to a linear time-dependent AR(p) model written in DLM form.

Consequently, B-splines DBSTAR(r, k, p) models are specified by rewriting the model from equation (4.26) into the observational process Y_t as in (4.18) conditional on the parameter vector $\underline{\theta}_t$ as in equation (4.19). The components of Δ_t are, hence-forward, $\underline{F}_t' = [\underline{z}_t', \underline{z}_t' B_1(t), \ldots, \underline{z}_t' B_m(t)]$ with a known $(m+1)(p+1)$-dimensional vector of polynomial regression variables, $\mathbf{G}_t$, $\mathbf{W}_t$ and V_t are defined as in Section

4.2.

Example 4.9. *For each time t, consider a simple B-splines DBSTAR$(3,1,1)$ model with no intercept, B-splines degree $r = 3$, $k = 1$ knot, autoregressive order $p = 1$ and transition variable $s_t = y_{t-1}$. The components are hence established as a 6-dimensional vector $\underline{F}_t' = [y_{t-1}, y_{t-1}B_1(t), y_{t-1}B_2(t), y_{t-1}B_3(t), y_{t-1}B_4(t), y_{t-1}B_5(t)]$, a 6×6 evolution matrix $\mathbf{G}_t = \mathbf{I}$, the identity matrix and a 6×6 state covariance matrix $\mathbf{W}_t$. Also, $\underline{\theta}_t = [\theta_{10t}, \theta_{21t}, \theta_{22t}, \theta_{23t}, \theta_{24t}, \theta_{25t}]$ is a 6-dimensional state vector containing parameters associated with each element of $\underline{F}_t$.*

4.4 Discussion

In this chapter, we described both the Taylor and Splines DBSTAR models for non-stationary nonlinear processes as alternative to both the classical and computational Bayesian STAR models. DBSTAR models belong to the class of nonlinear autoregressive models and are required for generating the data with the appropriate nonlinear characteristics in order to represent the underlying process.

DBSTAR models sequentially update their dynamic parameters as well as the observational variance through time in a Bayesian analytical fashion via Kalman filtering based on the Dynamic Linear Model formulations of West and Harrison (1997). Computationally cheap, sequential analytical forms for posterior parametric and forecast distributions are obtained at each time step that includes estimation of the interpretable parameters (by solving systems of equations).

DBSTAR models in their simplest forms can be seen as approximations of the classical STAR model, where the smooth transition function, such as the logistic in

(2.13), is represented by approximation methods, such as polynomial approximations or splines. We adopt a dynamic smoothing transition function $\pi(s_t; \gamma_t, c_t)$ where parameters γ_t and c_t, treated as unknown, are allowed to vary in time and to adapt to data sequentially. The transition variable s_t can be either an exogenous variable or a past value of the process, depending on the application. In the latter case, $s_t = y_{t-d}$ with delay parameter d determined a priori. Again, in cases where initial data are available, a model predictive performance approach, as described in 3.3, is straightforward for determining the unknown delay parameter d.

Some parameters in the DBSTAR model do not belong to the state vector for a few important reasons: (i) they do not need to be updated sequentially in time, (ii) analytical computations would be intractable, therefore we would lose the conjugacy properties in the prior-to-posterior updating analysis, (iii) they do not belong to the model definition originally. Bayesian model selection methods are therefore recommended for the following parameters, in case values cannot be specified beforehand:

- The Taylor series degree r, where $r = 1, 2, ..., r_{max}$,
- The Splines degree r, where $r = 1, 2, ..., r_{max}$,
- The autoregressive order p, where $p = 1, 2, ..., p_{max}$,
- The delay parameter d of the transition variable $s_t = y_{t-d}$, where $d = 1, 2, ..., d_{max}$,
- The number of knots k, where $k = 1, 2, ..., k_{max}$,
- The discount factor δ_W of the state variance $\mathbf{W}_t$, where $0 < \delta_W \leq 1$,
- The discount factor δ_V of the observational variance V_t, where $0 < \delta_V \leq 1$.

If all the above mentioned parameters can be individually identified, which is unlikely, this stage should be skipped and the parametric prior-to-posterior updating can be carried out using the given values. In case the modeller needs to decide about two possible values for one parameter, say for example, $p_{max} = 2$, so two candidate models will be under analysis, one for $p = 1$ and another for $p = 2$, then, the Bayes' factor described in section 3.3.1 should be used to identify p. In the case there is, however, a need of identifying more than two possible values for one or more parameters, then more than two candidate models will be under analysis, therefore the joint log-likelihood predictive should be instead used to select them.

Unlike the classical STAR and CBSTAR model formulations, the observational variance V_t of the underlying nonlinear AR process, Y_t, is treated as an unknown dynamic parameter whose distribution sequentially adapts to data in the usual normal-inverse-gamma conjugate analysis.

Chapter 5

Modelling further components in DBSTAR models

In this chapter we present extensions of DBSTAR models defined in the previous Chapter 4 and incorporate seasonal, cyclical, regression and other components into the model structure.

There are many time series processes, in practice, that present not only nonlinearity but also other rather important characteristics, such as global non-stationarity, seasonality or dependence on covariates. The existing class of STAR models might not deal easily with such processes.

This chapter specifies models that can capture these different features of time series by using components. Each of these different features is represented individually by a DLM component added together to the DBSTAR formulation, producing more general DBSTAR models for the processes under investigation. The components are appropriately accommodated by using additive decomposition techniques

(West and Harrison, 1997). Notice that it is also possible to specify the models using multiplicative decomposition. Alternatively, additive decomposition can be used for series for which multiplicative decomposition is more appropriate after a log-transformation. A DBSTAR formulation for modelling processes with multiple regimes is also described.

5.1 DBSTAR models for non-stationary processes

DBSTAR models, as defined in the previous Chapter 4, are suitable for modelling local non-stationary but global stationary processes free from additional model components. Recall from Section (2.1.5) that a locally non-stationary process is identified by splitting the time axis into small time intervals and checking whether non-stationarity is detected in each of them, and a globally non-stationary process is identified by checking the presence of non-stationarity using the whole period $t = 1, \ldots, T$.

Local non-stationarity is modelled by DBSTAR models since the parameters are time-dependent and, according to the observed data arriving sequentially over time, the local (upward or downward) trend can be incorporated into the dynamic parameters, at each time t. However, there are other processes that present global non-stationarity in their structures and the existing models described in Chapter 2 may not be appropriate to address those features. For instance, in Lopes and Salazar (2005), it was necessary to difference $(Y_t - Y_{t-1})$ the US IPI (Industrial Production Index in the US) data four times to achieve stationarity and remove possible seasonality before modelling it using the CBSTAR model. Many other

formulations from both classical and Bayesian inferences with the use of MCMC need to first difference the series then model the differenced series as the dynamic changes cannot be handled by stationary models so they eliminate the trend by differencing the series.

One rather important advantage of the formulation proposed in this section over existing models reviewed in Chapter 2 is that there is no need for differencing the series in case it presents non-stationarity. Our dynamic stochastic formulation proposed here can be directly applied to the undifferenced time series. Furthermore, either linear or nonlinear time trend curves can be accommodated into our formulation.

We define in this section the DBSTAR formulation for modelling both local and global non-stationarities as well as nonlinearity that may be conjointly present in the process. The difference between these and models from the previous Chapter 4 lies simply in an augmentation of elements in the quadruple Δ_t to accommodate components directly related to the presence of trend in the process. The addition of explicit components increases the dimension of vectors and matrices of the models structure compared with the previous formulation. This approach follows similar work for the DLM by West and Harrison (1997).

In this approach, the quadruple $\Delta_t = \{\underline{F}_t, \mathbf{G}_t, V_t, \mathbf{W}_t\}$ is extended to $\underline{F}_t = (\underline{F}_{1t}, \underline{F}_{2t})$, $\mathbf{G}_t = (\mathbf{G}_{1t}, \mathbf{G}_{2t})$ and $\mathbf{W}_t = (\mathbf{W}_{1t}, \mathbf{W}_{2t})$, where $\underline{F}_{1t}$, $\mathbf{G}_{1t}$ and $\mathbf{W}_{1t}$ are associated with the nonlinear autoregressive components as in Section 4.1, and $\underline{F}_{2t}$, $\mathbf{G}_{2t}$ and $\mathbf{W}_{2t}$ are associated with the trend components. DBSTAR models for non-

stationary processes are defined as follows,

$$(Y_t \mid \underline{\theta}_t, \underline{\tau}_t) \sim N\left(\underline{F}'_{1t}\underline{\theta}_t + \underline{F}'_{2t}\underline{\tau}_t, V_t\right) \tag{5.1}$$

$$(\underline{\theta}_t \mid \underline{\theta}_{t-1}) \sim T_{n_{t-1}}\left(\mathbf{G}_{1t}\underline{\theta}_{t-1}, \mathbf{W}_{1t}\right) \tag{5.2}$$

$$(\underline{\tau}_t \mid \underline{\tau}_{t-1}) \sim T_{n_{t-1}}\left(\mathbf{G}_{2t}\underline{\tau}_{t-1}, \mathbf{W}_{2t}\right) \tag{5.3}$$

where, $\underline{\theta}_t$ is defined as in Section 4.2 and $\underline{\tau}_t$ is a l-dimensional vector with the trend function defined as a l-th polynomial DLM. For instance, models with $l = 1$ have a component called the local level and $l = 2$ is referred to as the local level plus local linear growth rate. It is scarcely recommended $l > 3$, as the first two degrees are usually sufficient to represent most of the non-stationary behaviour of processes.

The l-dimensional vector $\underline{F}_{2t} = [1, 0, \ldots, 0]$ is a fixed canonical vector associated with the trend in $\underline{\tau}_t$, with 1 representing changes in the mean response, which gives a shift in the current level of the series and 0 representing the other elements of the state vector associated with the changes linearly through time, which may themselves also evolve. The evolution matrix of the trend components is as

$$\mathbf{G}_{2t} = \begin{bmatrix} 1 & 1 & 0 & 0 & \ldots & 0 \\ 0 & 1 & 1 & 0 & \ldots & 0 \\ \vdots & & \ldots & & \ddots & \vdots \\ 0 & 0 & \ldots & 0 & 1 & 1 \\ 0 & 0 & \ldots & 0 & 0 & 1 \end{bmatrix}$$

and the state covariance matrix $\mathbf{W}_{2t} = \text{diag}[\mathbf{W}_{2t}, \ldots, \mathbf{W}_{(l+1)t}]$ for which a discount factor δ_W, satisfying the condition $0 < \delta_W \leq 1$, is considered.

These DBSTAR models are thus more flexible than the previous version of the DBSTAR formulation when only a local linear trend could be modelled.

5.2 DBSTAR models for periodic processes

In this section we present DBSTAR models for modelling observed repetitive behaviour in terms of cyclical/seasonal components added to the model structure of Section 4.1. Notice that they can be also added to the model defined in the previous Section 5.1 in case the process presents both non-stationarity and cyclical patterns, including seasonality. In this thesis, we distinguish cycle and seasonality. The former has repetitive form within any period of time, such as hourly, daily, weekly, and so forth whereas the latter has variation within one calendar year, such as monthly, quarterly, etc.

A useful and functional representation of periodic forms is given in terms of trigonometric functions (West and Harrison, 1997, p.246). Fourier form representations are considered as sine/cosine waves which provide economic characterisation on parameters.

DBSTAR models for periodic processes extend the quadruple Δ_t to $\underline{F}_t = (\underline{F}_{1t}, \underline{F}_{2t})$, $\mathbf{G}_t = (\mathbf{G}_{1t}, \mathbf{G}_{2t})$ and $\mathbf{W}_t = (\mathbf{W}_{1t}, \mathbf{W}_{2t})$, where $\underline{F}_{1t}, \mathbf{G}_{1t}$ and $\mathbf{W}_{1t}$ are associated with the nonlinear autoregressive components as in Section 4.1, and $\underline{F}_{2t}, \mathbf{G}_{2t}$ and $\mathbf{W}_{2t}$ are associated with the periodic components. Hence, with explicit components for repeated patterns, DBSTAR models with h harmonics for periodic processes are defined as follows

$$\left(Y_t \mid \underline{\theta}_t, \underline{\psi}_t\right) \sim N\left(\underline{F}'_{1t}\underline{\theta}_t + \underline{F}'_{2t}\underline{\psi}_t, V_t\right) \tag{5.4}$$

$$(\underline{\theta}_t \mid \underline{\theta}_{t-1}) \sim T_{n_{t-1}} \left(\mathbf{G}_{1t}\underline{\theta}_{t-1}, \mathbf{W}_{1t}\right) \tag{5.5}$$

$$\left(\underline{\psi}_t \mid \underline{\psi}_{t-1}\right) \sim T_{n_{t-1}} \left(\mathbf{G}_{2t}\underline{\psi}_{t-1}, \mathbf{W}_{2t}\right) \tag{5.6}$$

where, $\underline{\psi}_t$ is a $2h$-dimensional vector with the periodic functions, which is a linear combination of trigonometric terms, $\underline{\psi}_t = [\sum_{j=1}^{h} S_j(t), \sum_{j=1}^{h} S_j^*(t)]$, i.e., a summation of the h harmonics, $S_j(t) = a_j cos(t\omega_j) + b_j sin(t\omega_j)$ and their conjugate $S_j^*(t) = -a_j sin(t\omega_j) + b_j cos(t\omega_j)$. The frequency of each harmonic is defined as $\omega_j = \frac{2\pi j}{s}$, with s the period of the cycle, for example, $s = 24$ for hourly data within a day or $s = 7$ for daily data within a week; furthermore, the seasonality, normally $s = 12$ for monthly data or $s = 4$ for quarterly data. The quantities a_j and b_j are called the Fourier coefficients.

The $2h$-dimensional vector $\underline{F}_{2t}$ is a canonical vector associated with the harmonics in $\underline{\psi}_t$, with 1 regarding the harmonic positions and 0 otherwise, which gives $\underline{F}_{2t} = [1, 0]$ for 1 harmonic, $\underline{F}_{2t} = [1, 0, 1, 0]$ for 2 harmonics, and so forth. The evolution matrix of the periodic components $\mathbf{G}_{2t}$ has $|\mathbf{G}_{2t}| = cos^2(t\omega_j) + sin^2(t\omega_j) = 1$. The discrete-time evolution of the j$-th$ harmonic from time t to time $t+1$ is given by

$$\begin{pmatrix} S_j(t+1) \\ S_j^*(t+1) \end{pmatrix} = \mathbf{G}_{2t} \begin{pmatrix} S_j(t) \\ S_j^*(t) \end{pmatrix}$$

The $(2h \times 2h)$-matrix $\mathbf{W}_{2t}$ contains the covariances of the cyclical components.

The first harmonic, called the fundamental harmonic, is expected to dominate the seasonal pattern, having a strong sinusoidal signal. The higher frequency harmonics oscillate faster than the fundamental one and are more appropriate for modelling higher frequency repetitive behaviour. Obviously, the larger the h the more accurate

the modelling of periodic variations in the data. However, adopting the parsimony principle we look for the smallest h that can still provide a good representation of the seasonality in the underlying process.

For cases where large enough initial data points are available to enable the investigation of the seasonal behaviour, the parsimonious value of h can be determined, for example, by applying a Bayesian approach, similar to those described in Section 3.3.

5.3 Incorporating predictor variables in DBSTAR models

It is straightforward to extend DBSTAR models to allow for predictor variables $\underline{x}_t' = [x_{1t}, \ldots, x_{qt}]$ over time as exogenous regressors to be incorporated into the formulations defined in Chapter 4 as well as in the previous sections 5.1 and 5.2.

This approach, called DBSTARX (eXogenous) models, investigates the dependence of a variable Y_t not only on the past values of the series but also on the values of other time series predictor variables. Therefore, autoregression and regression are both investigated in this formulation.

There are various possible sources of information that could be used as predictor variables in DBSTARX models. The vector $\underline{x}_t$ may accommodate (i) exogenous time series observed at same time points as the dependent variable Y_t, (ii) lagged exogenous time series, i.e., past values of other time series variables ($\underline{x}_{t-d}$, where d is the delay parameter) and/or (iii) dummy variables. Any number of predictor

variables can be added to the model structure.

For each time t, with p as the autoregressive order and q as the number of external variables, DBSTARX models extend the dimensions of the components in the quadruple Δ_t to a known $(r+1)(p+q+1)$-dimensional vector $\underline{F}_t$, a known $(r+1)(p+q+1) \times (r+1)(p+q+1)$ matrix $\mathbf{G}_t$, an unknown scalar V_t and an unknown $(r+1)(p+q+1) \times (r+1)(p+q+1)$ matrix $\mathbf{W}_t$. The components in Δ_t are associated with both the nonlinear autoregressive part of the model and the predictor variables, which uniquely define DBSTARX(q,r,p) models.

Firstly, with the addition of q exogenous variables, the model in equation (4.2) becomes as follows,

$$Y_t = \underline{z}_t'\underline{\phi}_{1t} + \underline{z}_t'\underline{\phi}_{2t}\pi(s_t;\gamma_t,c_t) + \underline{x}_t'\underline{\beta}_t + \epsilon_t \quad ; \quad \epsilon_t \sim N(0,\sigma_t^2) \tag{5.7}$$

Using the components of Δ_t as well as a dynamic transition nonlinear function $\pi(s_t;\gamma_t,c_t)$ with real values in the interval $[0,1]$, DBSTARX(q,r,p) models are defined by rewriting the model in the equation (5.7) into the observational process Y_t conditional on the parameter vector $\underline{\theta}_t$ in (4.18), where $\underline{F}_t' = [\underline{z}_t', \underline{z}_t's_t, \ldots, \underline{z}_t's_t^r, \underline{x}_t']$ with a known $(r+1)(p+q+1)$-dimensional vector of polynomial regression variables $s_t^i\underline{z}_t'$ $(i = 0,1,\ldots,r)$ with $\underline{z}_t' = (1, y_{t-1}, \ldots, y_{t-p})$ and $\underline{x}_t' = (x_{1t}, \ldots, x_{qt})$; $\underline{\theta}_t$ is the state vector containing parameters associated with the components of $\underline{F}_t$, i.e., $\underline{\theta}_t' = (\underline{\theta}_{1t}, \ldots, \underline{\theta}_{(r+1)t}, \underline{\beta}_t)$. Notice that all the parameters in $\underline{\theta}_t$ may vary in time, including the $\underline{\beta}_t$'s.

The $(r+1)(p+q+1)$-dimensional state vector $\underline{\theta}_t$ has a conditional multivariate Student-t distribution with n_{t-1} degrees of freedom as in (4.19), with $\mathbf{G}_t$ as the

state evolution matrix with elements g_{ijt} (for row i and column j) chosen according to the desired structural form of association between $\underline{\theta}_t$ and $\underline{\theta}_{t-1}$. In the case that no structural form is known, the random walk is used by setting $\mathbf{G}_t = \mathbf{I}$, the identity matrix. Furthermore, $\mathbf{W}_t$ is the state covariance matrix, for which a discount factor δ_W, satisfying the condition $0 < \delta_W \leq 1$, is considered as in (4.2) given by

$$\mathbf{W}_t = \left(\frac{1-\delta_W}{\delta_W}\right) \mathbf{G}_t \mathbf{C}_{t-1} \mathbf{G}_t'.$$

where $\mathbf{C}_{t-1}$ is a prior covariance matrix for $\underline{\theta}_t$. Notice that in the case when $\delta_W = 1$, the parameters have zero variance ($\mathbf{W}_t = \mathbf{0}$).

Example 5.1. *At each time t, a simple Taylor DBSTARX$(1,3,1)$ model with $q = 1$ external variable (x_{1t}) as well as a Taylor series expansion of degree $r = 3$, autoregressive order $p = 1$ (with no intercept) and transition variable $s_t = y_{t-1}$ is specified by the distribution of the observational process in equation (4.18) and the distribution of the state vector in equation (4.19), with the components of the quadruple Δ_t as a 5-dimensional vector $\underline{F}_t' = [y_{t-1}, y_{t-1}^2, y_{t-1}^3, y_{t-1}^4, x_{1t}]$, a 5×5 evolution matrix $\mathbf{G}_t$ and a 5×5 state covariance matrix $\mathbf{W}_t$. Also, $\underline{\theta}_t = [\theta_{0t}, \theta_{1t}, \theta_{2t}, \theta_{3t}, \beta_{1t}]$ is a 5-dimensional state vector containing parameters associated with each element of $\underline{F}_t$. Notice that in case θ_{1t}, θ_{2t} and θ_{3t} are concomitantly zero, the Taylor DBSTARX(1,3,1) model is reduced to a linear ARX(1) model written in DLM form.*

5.4 Multiple Regimes DBSTAR models

DBSTAR models can be extended to a more embracing form in order to accommodate more than 2 regimes. We refer to this extension as multiple regimes DBSTAR models which present $m = 2^k$, $k > 1$ regimes.

Like classical multiple regimes STAR models by van Dijk and Franses (1999) defined in equation (2.17), multiple regimes DBSTAR models also have $m-1$ transition variables as well as $m-1$ smoothness and threshold parameters to compose m regime models.

Generic multiple regimes DBSTAR models are thus defined based upon the use of approximation methods to all dynamic transition functions $\pi_j(s_{jt}; \gamma_{jt}, c_{jt}), j = 1, \ldots, m-1$, with nonlinear parameters γ_{jt} and $c_{jt}]$, as follows.

$$y_t \cong \underline{z}_t'\underline{\phi}_{1t} + \underline{z}_t'\underline{\phi}_{2t}\tilde{\pi}_1(s_{1t}; \gamma_{1t}, c_{1t}) + \ldots + \underline{z}_t'\underline{\phi}_{mt}\tilde{\pi}_{(m-1)}(s_{(m-1)t}; \gamma_{(m-1)t}, c_{(m-1)t}) + \xi_t \quad (5.8)$$

where $\tilde{\pi}_j(s_{jt}; \gamma_{jt}, c_{jt}), j = 1, \ldots, m-1$, are the approximations for the dynamic transition functions in the range $[0, 1]$ and $\xi_t \sim N(0, \sigma_t^2)$.

To define multiple regimes DBSTAR models, with m regimes, using a Taylor series expansion of degree r and with p as autoregressive order, we first need to approximate each of the dynamic nonlinear transition functions $\tilde{\pi}_j$ with the most appropriate Taylor series expansion as in equation (3.10), so the models as in equa-

tion (4.4) become as follows,

$$\begin{aligned} y_t \cong\ & \underline{z}_t'\underline{\theta}_{01t} + \left[(\underline{z}_t' s_{1t})\underline{\theta}_{11t} + (\underline{z}_t' s_{1t}^2)\underline{\theta}_{12t} + \ldots + (\underline{z}_t' s_{1t}^r)\underline{\theta}_{1rt}\right] + \ldots + \\ & + \left[(\underline{z}_t' s_{(m-1)t})\underline{\theta}_{m1t} + (\underline{z}_t' s_{(m-1)t}^2)\underline{\theta}_{m2t} + \ldots + (\underline{z}_t' s_{(m-1)t}^r)\underline{\theta}_{mrt}\right] + \xi_t \end{aligned} \quad (5.9)$$

where $\xi_t \sim N(0, \sigma_t^2)$ and for $i = 1, \ldots, r$ and $j = 1, \ldots, m$, $\underline{\theta}_{jit} = \beta_{jit}(\gamma_{jt}, c_{jt})\underline{\phi}_{2t}$, with $\beta_{jit}(\gamma_{jt}, c_{jt})$ are polynomial functions of γ_{jt} and c_t obtained from the i-th coefficient of the Taylor series expansion within the j-th regime.

Similarly to define DBSTAR models, for each time t, the quadruple $\Delta_t = \{\underline{F}_t, \mathbf{G}_t, V_t, \mathbf{W}_t\}$ contains vector and matrices with higher dimensions, as $\underline{F}_t$ is a known $(m-1)(r+1)(p+1)$-dimensional vector, $\mathbf{G}_t$ is a known $(m-1)(r+1)(p+1) \times (m-1)(r+1)(p+1)$ matrix, V_t is an unknown scalar and $\mathbf{W}_t$ is an unknown $(m-1)(r+1)(p+1) \times (m-1)(r+1)(p+1)$ matrix.

Using the components of Δ_t as well as dynamic transition nonlinear functions $\pi_j(s_{jt}; \gamma_{jt}, c_{jt})$ with real values in the interval $[0, 1]$, Taylor multiple regimes DBSTAR models are defined by rewriting the model in equation (5.9) into the observational process Y_t conditional on the parameter vector $\underline{\theta}_t$ as

$$(Y_t \mid \underline{\theta}_t) \sim N\left(\underline{F}_t'\underline{\theta}_t, V_t\right) \quad (5.10)$$

where $\underline{F}_t' = [\underline{z}_t', \underline{z}_t' s_{1t}, \ldots, \underline{z}_t' s_{1t}^r, \ldots \underline{z}_t' s_{mt}, \ldots, \underline{z}_t' s_{mt}^r]$ with a known $(m-1)(r+1)(p+1)$-dimensional vector of polynomial regression variables $s_{jt}^i \underline{z}_t'$ $(i = 0, 1, \ldots, r, j = 0, 1, \ldots, m)$ with $\underline{z}_t' = (1, y_{t-1}, \ldots, y_{t-p})$; $\underline{\theta}_t$ is the state vector containing parameters associated with the components of $\underline{F}_t$, i.e, $\underline{\theta}_t' = (\underline{\theta}_{01t}, \underline{\theta}_{11t}, \ldots, \underline{\theta}_{1rt}, \underline{\theta}_{m1t}, \ldots, \underline{\theta}_{mrt})$.

The $(m-1)(r+1)(p+1)$-dimensional state vector $\underline{\theta}_t$ has a conditional multivariate Student-t distribution with n_{t-1} degrees of freedom

$$(\underline{\theta}_t \mid \underline{\theta}_{t-1}) \sim T_{n_{t-1}} (\mathbf{G}_t \underline{\theta}_{t-1}, \mathbf{W}_t) \tag{5.11}$$

where $\mathbf{G}_t$ is the state evolution matrix according to the desired structural form of association between $\underline{\theta}_t$ and $\underline{\theta}_{t-1}$, and $\mathbf{W}_t$ is the evolution covariance matrix, both are specified as in Section 4.2.

5.5 General DBSTAR models

Each approach described individually in the previous sections of this chapter may be combined together to form a generic formulation which we refer to as General DBSTAR models.

The clearest way to examine a time series and decide which components should be included in the model is with exploratory analysis, such as scatterplots, boxplots, autocorrelation plots, partial autocorrelation plots, and other techniques. Those analyses would support the separation of each characteristic into components representing trend, seasonality, cyclical irregularity, and so forth.

General DBSTAR models extend the quadruple Δ_t to $\underline{F}_t = (\underline{F}_{1t}, \ldots, \underline{F}_{kt})$, $\mathbf{G}_t = (\mathbf{G}_{1t}, \ldots, \mathbf{G}_{kt})$ and $\mathbf{W}_t = (\mathbf{W}_{1t}, \ldots, \mathbf{W}_{kt})$, where k is the number of desirable components, $\underline{F}_{1t}$, $\mathbf{G}_{1t}$ and $\mathbf{W}_{1t}$ are associated with the nonlinear autoregressive components as in Section 4.1, and $(\underline{F}_{2t}, \ldots, \underline{F}_{kt})$, $(\mathbf{G}_{2t}, \ldots, \mathbf{G}_{kt})$ and $(\mathbf{W}_{2t}, \ldots, \mathbf{W}_{kt})$ are associated with the additional components.

General DBSTAR models are defined as follows,

$$(Y_t \mid \underline{\theta}_{1t} \dots \underline{\theta}_{kt}) \sim N\left(\underline{F}'_{1t}\underline{\theta}_{1t} + \dots + \underline{F}'_{kt}\underline{\theta}_{kt}, V_t\right) \tag{5.12}$$

$$(\underline{\theta}_{1t} \mid \underline{\theta}_{1t-1}) \sim T_{n_{t-1}}\left(\mathbf{G}_{1t}\underline{\theta}_{1t-1}, \mathbf{W}_{1t}\right) \tag{5.13}$$

$$\dots$$

$$(\underline{\theta}_{kt} \mid \underline{\theta}_{kt-1}) \sim T_{n_{t-1}}\left(\mathbf{G}_{kt}\underline{\theta}_{kt-1}, \mathbf{W}_{kt}\right) \tag{5.14}$$

where, $\underline{\theta}_{1t}$ is a vector with the parameters associated with the nonlinear part of the model and $\underline{\theta}_{2t}, \dots, \underline{\theta}_{kt}$ contain the parameters associated with the additional components. The matrix $\mathbf{G}_t$ is the state evolution matrix with elements g_{ijt} (for row i and column j) chosen according to the desired structural form of association between $\underline{\theta}_t$ and $\underline{\theta}_{t-1}$. Furthermore, $\mathbf{W}_t$ is the state covariance matrix, for which a discount factor δ_W, satisfying the condition $0 < \delta_W \leq 1$, is considered as in (4.2) given by

$$\mathbf{W}_t = \left(\frac{1-\delta_W}{\delta_W}\right)\mathbf{G}_t\mathbf{C}_{t-1}\mathbf{G}'_t.$$

where $\mathbf{C}_{t-1}$ is a prior covariance matrix for $\underline{\theta}_t$. Notice that in the case when $\delta_W = 1$, the parameters have zero variance ($\mathbf{W}_t = \mathbf{0}$).

5.6 Discussion

In this chapter, we have introduced a General DBSTAR model along with interesting special cases. The addition of further components in DBSTAR models are convenient when working with general time series data and are suitable for modelling a wide

variety of behaviours, such as, trend, cycle or seasonality.

General DBSTAR models as proposed in Section 5.5 can be interpreted as decomposing an observed time series into various components. These, however, are not the most general possible (see Chapter 8 for extensions), but they do encompass a wide variety of models. To understand the types of behaviour General DBSTAR models allow for, it is helpful to discuss their particular cases. In chapter 4, we proposed DBSTAR models suitable for local non-stationary processes, in which the coefficients in the state vector $\underline{\theta}_t$ may accommodate a short-term stochastic trend, being permissible to change only locally over time. However, in general, the trend behaviour implies that a time series can wander expansively. DBSTAR models for non-stationary processes proposed in Section 5.1 have further components to accommodate the long-term stochastic trend terms. Those components are responsible for measuring the relative sizes of the trend and may have variance which may change over time increasing or decreasing, therefore probability distributions were given at each time t.

Cycles and seasonality can be modelled with DBSTAR models proposed in Section 5.2. Classical and computational Bayesian STAR models use high autoregressive order trying to capture the periodic behaviour. On the other hand, DBSTAR models for periodic processes use low autoregressive order and model the cycle or seasonality explicitly using Fourier analyses. By doing so, the models are more parsimonious and more than that, they give accurate specification of the cyclical/seasonal behaviour of the time series which are key features for modelling such processes.

Even though the use of external variables may be in a nonlinear form, such as

the variable s_t in $\pi(s_t;.)$, DBSTARX models were proposed for processes for which there is still some additional linear relationships between exogenous and dependent variables. In this approach, both autoregression and regression are accommodated.

DBSTAR models trivially deal with missing data. Other existing approaches, in particular using STAR type models, have enormous difficulties in handling missing values (West and Harrison, 1997). Recall that $D_t = (y_t, D_{t-1})$ represents the information available at time t after observing Y_t. So, for the case of a missing value y_t, the information set at time t is just the previous available information, i.e., $D_t = D_{t-1}$. On the one hand, during the filtering stage of the parametric prior-to-posterior updating procedure, there is no new information to be incorporated so the prior probability distributions of each $\underline{\theta}_{it}, i = 1, \ldots, k$, from model (5.12)-(5.14), i.e., $(\underline{\theta}_{it} \mid D_{t-1}) \sim T_{n_{t-1}}(\underline{a}_{it}, S_{t-1}\mathbf{R}_{it})$ are not updated, therefore the posterior distributions are equal to the prior distributions, viz., $(\underline{\theta}_{it} \mid D_t) \sim T_{n_t}(\underline{m}_{it}, \mathbf{C}_{it})$, with $\underline{m}_{it} = \underline{a}_{it}$ and $\mathbf{C}_{it} = S_t\mathbf{R}_{it}$ as well as $S_t = S_{t-1}$ and $n_t = n_{t-1}$. This sequential algorithm will then update the distributions once a new information becomes available (West and Harrison, 1997). Nevertheless, if a new observation does not become available, forecast distributions can still be obtained. On the other hand, it should be pointed out that uncertainty should really increase as a consequence of missing data. Using this sequential algorithm, this does not actually happen. For a missing value y_t, the posterior distribution $(\underline{\theta}_{it} \mid D_t) \sim T_{n_t}(\underline{m}_{it}, \mathbf{C}_{it})$ should not have posterior variance equals to the prior variance $(\mathbf{C}_{it} = S_t\mathbf{R}_{it})$ but instead an increase in the variance should be taken into account, such as $\mathbf{C}_{it} = S_t\mathbf{R}_{it} + \mathbf{U}_{it}$, where $\mathbf{U}_{it}$ represents uncertainty due to a missing data. By doing so, DBSTAR models are

also able to produce accurate forecast distributions since a 2-step ahead forecast distribution has greater uncertainty than a 1-step ahead forecast distribution.

Chapter 6

Modelling Canadian lynx data

In this chapter, appropriate formulations of DBSTAR models defined in chapters 4 and 5 are applied to the well-known Canadian lynx data set. Approaches described in both chapters can account for the cyclic behaviour observed in the data either through a high AR order or with the inclusion of components.

The aim here is twofold. On the one hand, the aim is to validate proposed DBSTAR models by comparing their fitting performances against the performances of both the classical STAR and CBSTAR models. On the other hand, the aim is to illustrate the extra features that can be achieved for modelling nonlinear time series by adopting sequential approaches with time-dependent parameters.

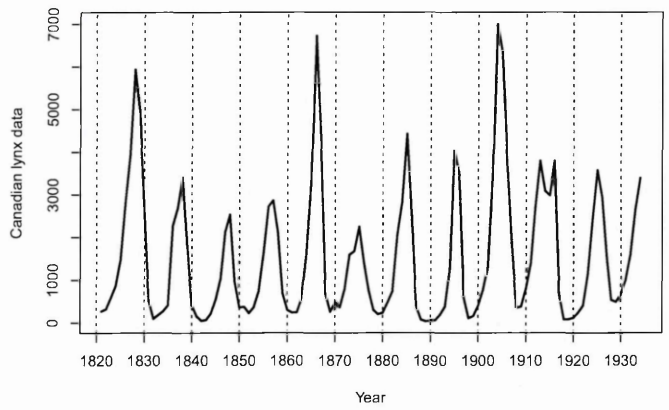
Section 6.1 describes the data set and exploratory analyses to demonstrate why DBSTAR models may be appropriate for modelling this series. Section 6.2 describes formulations of DBSTAR models suitable for the Canadian lynx series. Section 6.3 describes selected DBSTAR models based on the retrospective smoothing procedure adopted to allow goodness of fitting measures to be compared to existing static

STAR approaches. Section 6.4 presents results comparing the fitting performances of DBSTAR models with the classical STAR and CBSTAR models. The chapter ends in Section 6.6 with a discussion of this application exercise.

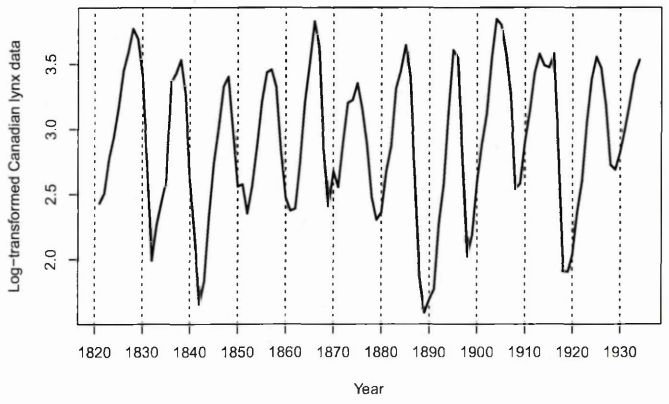
6.1 The data set and initial data analysis

The Canadian lynx data set consists of the annual series of the number of Canadian lynx trapped in the Mackenzie River district of North-west Canada from 1821 to 1934, giving therefore a total of 114 observations. This data set was first published by Elton and Nicholson (1942) and originally recorded to support the understanding of the population dynamics of the ecological system in that area. Analyses of this time series can be found, for example, in Tong (1990), Teräsvirta (1994) and Lopes and Salazar (2005).

Figure 6.1 (a) and (b) show the time series plots of the Canadian lynx data and the log-transformed series, respectively. As can be seen in both plots, the most prominent features of the Canadian lynx time series are (i) the lack of trend, (ii) the presence of irregular changes in the amplitude over time, and (iii) the presence of persistent non-regular cyclic oscillations with periods of 10 or 11 years. Regarding the cyclic oscillations, observe that there is an apparent asymmetry in each of the cycles with long ascent times occurring between 5 and 7 years followed by shorter descent times of only about 3 and 4 years. The vertical dashed lines at intervals of 10 years in both plots help to identify such irregularities. These features have been familiar to biologists for a long time and are prominent in historical records of trappings of lynx in Canada (see, for example, Elton and Nicholson (1942) and



(a)



(b)

Figure 6.1: (a) Original Canadian lynx time series and (b) log-transformed Canadian lynx time series, yearly observed from 1821 to 1934

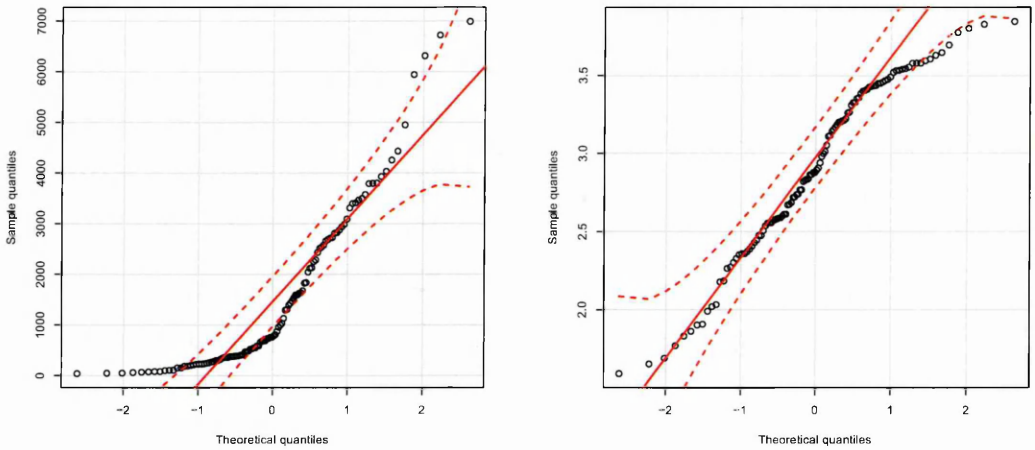
the references therein). The asymmetry in the cycles suggests the use of nonlinear models, since linear models fail to deal with those characteristics (Moran, 1953).

The original time series was $\log_{10}$-transformed to (i) remove the marked right-skewness of its frequency plot, (ii) bring some outliers close to the other data points, and (iii) allow for comparative fitting analysis, as both the classical STAR and CBSTAR models were applied to the log-transformed series. Figure 6.2 (a) and (b) show Normal Q-Q Plots of the Canadian lynx time series and the log-transformed

series, respectively, to illustrate the points (i) and (ii) just mentioned.

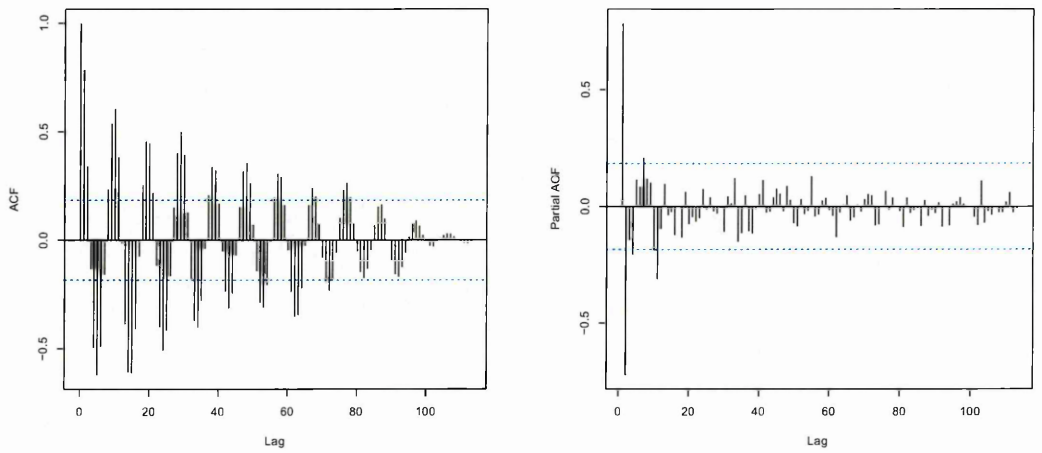
Although the Canadian lynx data are counts and, therefore, a more appropriate probability distribution to model them being the Poisson distribution, researchers have been using the normal distribution for the log-transformed data instead (including Tong (1990), Teräsvirta (1994) and Lopes and Salazar (2005)). In this thesis, the normal distribution for the log-transformed data is also considered. The Normal Q-Q plot of the original series in Figure 6.2 Panel (a) presents (i) the points following a persistent nonlinear pattern with great departures from the 45-degree reference straight line for larger negative and positive values of the theoretical quantiles, (ii)points not covered by the 99% confidence interval, being then considered outliers, and (iii) high level of positive skewness: all of those suggest that the data are not normally distributed. After taking the log-transformation, the Normal Q-Q plot of the log-transformed series in Panel (b) shows (i) most points fall approximately along the reference line, (ii) most points lying within the 99% confidence interval with only a few outliers at the high end of the line: these suggest that the data can be reasonably approximated by the normal distribution. Moreover, the reasons the data are log-transformed are also because DBSTAR models make the assumption of normality of the underlying process so that the Kalman filter can be implemented, and additionally, the performance from DBSTAR models will be compared to the performance of existing approaches, in which the data was also log-transformed.

Figure 6.3 (a) and (b) show the autocorrelation function (ACF) and the partial ACF (PACF) plots of the log-transformed series, respectively. There seems to be a



(a) (b)

Figure 6.2: Normal Q-Q Plot of (a) original Canadian lynx time series and (b) log-transformed series



(a) (b)

Figure 6.3: (a) Autocorrelation function (ACF) and (b) Partial ACF of log-transformed series

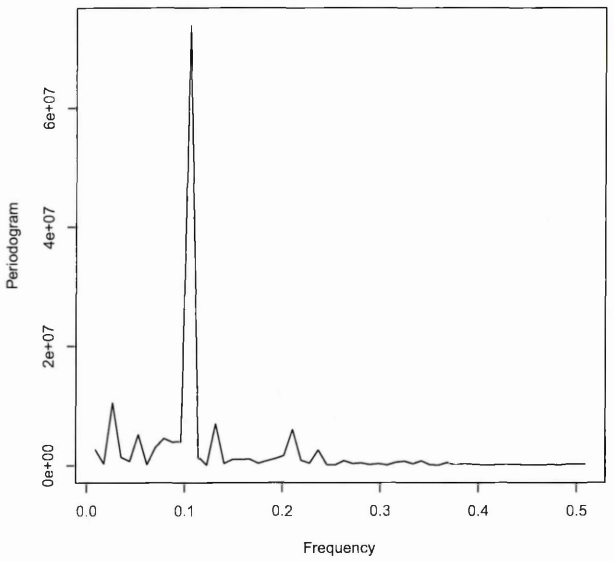


Figure 6.4: Periodogram of original Canadian lynx time series

slow sinusoidal decay along time in the ACF together with a cut after lag 2 in the PACF, assuming the PACFs on lags 4 and 7 are spurious. Such behaviour from both the ACF and PACF plots suggest that the Canadian lynx data should be modelled by an autoregressive model. Notice that there is a significant peak at lag 10 in the PACF that seems to show a cyclical period of 10 years.

Figure 6.4 shows the Bayesian periodogram (Bretthorst, 1988) of the Canadian lynx data set. The periodogram gives a better idea about the periodicity of the cycle in the data set, particularly since the cycles are not related to the regularly seasonality (when the periodic repetition occurs within one-calendar year), which would have given 12 periods for monthly data or 3 periods for quarterly data.

The Bayesian periodogram method uses the log-likelihood of a single sinusoidal regression model $y_t = acos(2\pi t/\lambda) + bsin(2\pi t/\lambda) + \epsilon_t$, where $\epsilon_t \sim N(0, \sigma^2)$ and λ is the periodicity or wavelength (the time it takes to complete a full cycle) of

the process Y_t, to determine the variation of the amplitude with the frequency. The predominant spike of the periodogram appears at around the 0.1 frequency. Exploration of the periodogram values points out that the peak occurs at nearly exactly this frequency. This indicates a wavelength of $1/0.10 = 10$ years, suggesting a sustained and persistent cyclical feature of that time period. Consequently, there appears to be a dominant periodicity of about 10 years in the Canadian lynx data set and harmonics with wavelength of 10 years should be included in DBSTAR models. This result was also obtained by Elton and Nicholson (1942), who found the length of 10 years for the cyclic repetition of the Canadian lynx data.

6.2 Formulating DBSTAR models

Various DBSTAR formulations were applied to the Canadian lynx time series, which are described in this section. Figure 6.5 summarises each particular version of DBSTAR models that was implemented. The diagram should be read from the top to the bottom. The first division relates to DBSTAR models using two alternative approximation methods to the nonlinear transition function: the Taylor series defined in Section 4.2 or Splines defined in Section 4.3. Notice that the latter could make use of either the Truncated Power Basis or the B-splines function. The subsequent partition reveals the standard formulations, free from additional model components described in Chapter 4, or the periodic approaches described in Section 5.2. The last splitting identifies the models depending on the specification of their state covariance matrices $\mathbf{W}_t$, which indicates whether the parameters in the state vectors are either static or dynamic, in the sense that they do or do not change over time,

respectively (if $\mathbf{W}_t = \mathbf{0}$, then the model is static). All possible combinations are then individually identified as Model 1, Model 2, ..., Model 12.

In this application, the Taylor DBSTAR(r, p) formulations defined in Section 4.2 use the Taylor series expansion of the logistic function in (2.13) as the smooth transition function between the two regimes. For this specification, the Taylor series degree, r, needs to be odd, since for even degrees the derivatives of the logistic function evaluated at the point we are expanding around are zero. For high Taylor series degree ($r = 5, 7, 9, \ldots$), approximations have only marginal improvements. Therefore, for parsimony reasons, we use $r = 3$ in the models. In addition, lagged values of the log-transformed series are used as the transition variable, so that $s_t = y_{t-d}$, as in Teräsvirta (1994) and Lopes and Salazar (2005).

Splines DBSTAR(r, k, p) models defined in Section 4.3 use either the Truncated Power Basis or the B-splines to represent the nonlinear transition function. In both approaches, both the polynomial degree r and the number k of knots need to be defined. In this application, we used the most common choice $r = 3$, which characterises the cubic splines, and k was decided by a model selection criterion. Again, lagged values of the log-transformed series are used as the transition variable, so that $s_t = y_{t-d}$.

There are other parameters which are not accommodated in the state vectors either and the use of the Bayesian approach based on the predictive performance of different models, described in Section 3.3, is recommended, as initial data are available. Those parameters were then considered unknown and we had to specify (i) the autoregressive order p, where $p = 1, 2, ..., 12$; (ii) the delay parameter d of

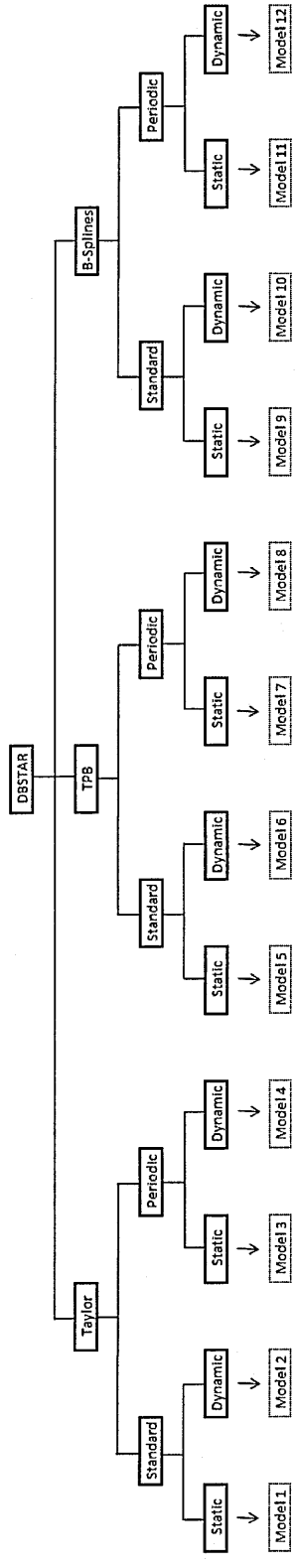


Figure 6.5: DBSTAR model configurations structured by a classification tree

the transition variable $s_t = y_{t-d}$, where $d = 1, 2, ..., 12$; (iii) the discount factor δ_W of the state variance $\mathbf{W}_t$, where $0 < \delta_W \leq 1$ ($\delta_V = 1$), (iv) the number of knots k and (v) the number of harmonics h.

Several models were run which differed only in values of these five parameters and those were selected by the log-smoothing likelihood (LSL) conditional on the parameters, as defined in Section 3.3.3. The LSL criterion gives evidence in favour of a model which presents the largest value of LSL. The reason the LSL criterion is used rather than the log-predictive likelihood (LPL) is that we aim to select dynamic models based on fitting performance rather than forecasting performance in order to make them comparable to both of the static approaches, the classical STAR and CBSTAR models. The use of the LPL criterion is shown in Chapter 7, where the focus is on selecting models for forecasting purposes.

This step is evaluated jointly with the parametric prior-to-posterior updating described in Section 3.2 using the above given values. The LSL criterion selected $h = 2$ for periodic DBSTAR models, with the components dominating the cyclical pattern by having strong sinusoidal signals.

Both the observational variance and the state variance matrices are specified to be the same for all models. The former is set as $V_t = V$, which means the unknown observational variance of the underlying nonlinear AR process was not detected to be stochastically changing from exploratory analyses, for this reason a variance discount technique for V_t was not used in this analysis. This implies we can make use of a simpler algorithm, since V_t is set to be fixed and does not have to be updated sequentially over time, and also the discount factor associated with it

does not have to be identified. So that, $\delta_V = 1$ and for the state covariance matrix $\mathbf{W}_t$, the use of its discount factor δ_W is needed in order to allow (slow) dynamic parameters adaptation.

The remaining two components in Δ_t, i.e., the polynomial regression variables from $\underline{F}_t$ and the evolution state matrix $\mathbf{G}_t$, vary according to the model specifications, as follows:

- Models 1 and 2: $\underline{F}_t^{'} = [\underline{z}_t^{'}, \underline{z}_t^{'}s_t, \underline{z}_t^{'}s_t^2, \underline{z}_t^{'}s_t^3]$ and $\mathbf{G}_t = \mathbf{I}$;
- Models 3 and 4: $\underline{F}_{1t}$ and $\mathbf{G}_{1t}$ are set as above for Models 1 and 2, and $\underline{F}_{2t}^{'} = [1, 0, 1, 0]$ and $\mathbf{G}_{2t}$ has $|\mathbf{G}_{2t}| = cos^2(t\omega_j) + sin^2(t\omega_j) = 1$, which gives prespecified trigonometric evolution changes for the state parameters from time $t-1$ to t;
- Models 5 and 6: $\underline{F}_t^{'} = [\underline{z}_t^{'}, \underline{z}_t^{'}s_t, \underline{z}_t^{'}s_t^2, \underline{z}_t^{'}s_t^3, z_t(s_t - \kappa_1)_+^3, \ldots, z_t(s_t - \kappa_k)_+^3]$ and $\mathbf{G}_t = \mathbf{I}$;
- Models 7 and 8: $\underline{F}_{1t}$ and $\mathbf{G}_{1t}$ are set as above for Models 5 and 6, and $\underline{F}_{2t}$ and $\mathbf{G}_{2t}$ are set as for Models 3 and 4;
- Models 9 and 10: $\underline{F}_t^{'} = [\underline{z}_t^{'}, \underline{z}_t^{'}B_1(t), \ldots, \underline{z}_t^{'}B_m(t)]$ and $\mathbf{G}_t = \mathbf{I}$; and
- Models 11 and 12: $\underline{F}_{1t}$ and $\mathbf{G}_{1t}$ are set as above for Models 9 and 10, and $\underline{F}_{2t}$ and $\mathbf{G}_{2t}$ are set as for Models 3 and 4.

Notice that when the evolution state matrix is set as $\mathbf{G}_t = \mathbf{I}$, the identity matrix, random walk evolution changes are assumed for the state parameters from time $t-1$ to t. Moreover, the state vector $\underline{\theta}_t$ accommodates the polynomial coefficients

associated with the polynomial regression variables in $\underline{F}_t$ and, for periodic models, $\underline{\psi}_t$ is a state vector associated with the linear combination of trigonometric terms for representing the periodic behaviour. Both vectors $\underline{\theta}_t$ and $\underline{\psi}_t$ are sequentially updated over time as shown in Section 3.2.2. Non-informative prior distributions were used to form initial relevant views about the future for all model parameters, as we did not have any prior knowledge about them. Thus, we set $\underline{m}_0 = \underline{0}$ and large variances $\mathbf{C}_0 = 3\mathbf{I}$ as the prior mean vector and covariance matrix, respectively.

6.3 Configuring DBSTAR models

The Kalman filtering together with the Kalman smoothing algorithms, were implemented using the software R (version 2.15.2). In order to obtain the results that account for the whole Canadian lynx data set at once, a retrospective analysis using Kalman smoothing was carried out on dynamic models. As with Kalman filtering, Kalman smoothing can be straightforwardly implemented as a backward-recursive algorithm, which depends only on the data used for filtering and one-step-ahead forecast moments. A retrospective analysis makes static and dynamic models comparable in the sense that it uses the whole data set after the Kalman filtering is applied, as described in Section 3.2.3. Firstly, we need the posterior probability distribution for $\underline{\theta}_t$, at each time t, provided by the Kalman filter. Then, the Kalman smoothing provides the conditional probability distributions of $\underline{\theta}_t$ given all the data D_T, for any time $t < T$.

Table 6.1 presents the final DBSTAR model configurations. This table relates to Figure 6.5 to name the type of each DBSTAR models under analysis. Both the

Taylor series and splines degrees were fixed as $r = 3$. Splitting the table in two parts, it is possible to reveal some important points: (i) the autoregressive order is higher for the standard approaches than for periodic models, (ii) the number of parameters to be updated sequentially in time is large, particularly so for the standard approaches, (iii) the delay parameter is selected as $d = 3$ for all model configurations.

A point to note from point (i) above is that standard models try to capture the cyclic behaviour with the use of large autoregressive orders. On the other hand, periodic models use low autoregressive orders and model the cycle explicitly using Fourier analyses. This gives an accurate estimate of the cyclical behaviour of the Canadian lynx data set which is a key feature for modelling such series. From (ii), the large number of parameters is directly related to the type of approximation methods as well as to the autoregressive orders. Notice that, in general, the number of parameters is lower for periodic models. Comparing the pairs of corresponding static models, which differ only in whether they are standard or periodic models, i.e., (1 and 3), (5 and 7) and (9 and 11), we note a reduction in the number of parameters by about 50%. For pairs of dynamic models, the reduction is of around 30% for models (2 and 4) and (6 and 8) and of 68% for the pair (10 and 12). For parsimony reasons, periodic models should be considered for further analyses.

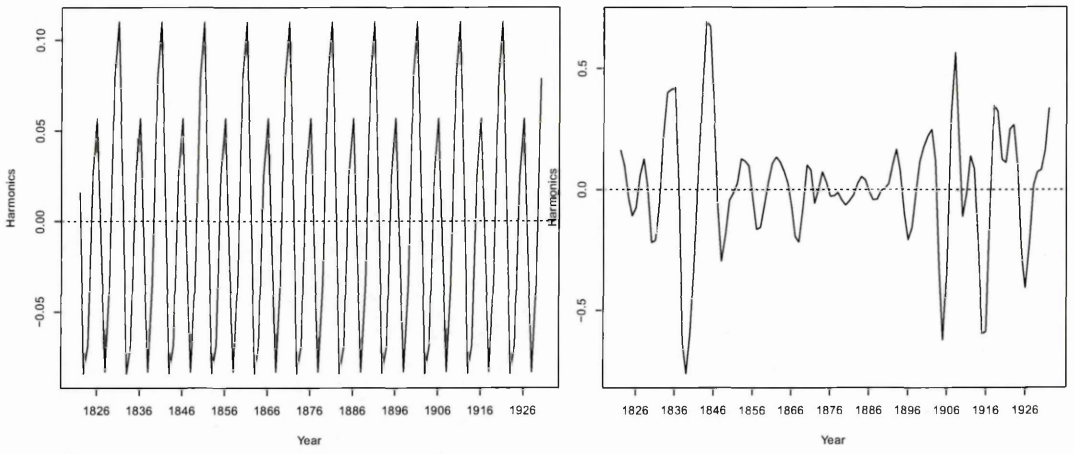
The LSL criterion selects $\delta_W < 1$, i.e., to model the Canadian lynx data we should make use of DBSTAR models with time-dependent parameters. Notice that the parameters of those models vary in time but slowly as the discount factor is less than 1 but not much, i.e., either $\delta_W = 0.80$ or $\delta_W = 0.85$. Nonetheless, DBSTAR models

Table 6.1: Final DBSTAR model configurations for the Canadian lynx data set

Type	Model	AR order	Polynomial degree	Delay	Harmonics	Knots	δ_W	No. param.
Standard	1	12	3	3	-	-	1	65
	2	9	3	3	-	-	0.85	50
	5	11	3	3	-	5	1	156
	6	8	3	3	-	5	0.85	108
	9	12	3	3	-	5	1	130
	10	8	3	3	-	5	0.85	108
Periodic	3	5	3	3	2	-	1	34
	4	5	3	3	2	-	0.85	34
	7	5	3	3	2	5	1	76
	8	5	3	3	2	5	0.85	76
	11	5	3	3	2	5	1	64
	12	2	3	3	2	5	0.80	34

with $\delta_W = 1$ were also analysed in order to have static versions of DBSTAR models to make them directly comparable to existing models, which are static approaches.

Figure 6.6 (a) and (b) show the evolution over time of the 2 harmonics of both periodic models 11 and 12, respectively. Recall that Model 11 does not allow the parameters to vary in time whilst Model 12 does. As can be seen in Panel (a), the harmonics for Model 11 are constant in time and there is a constant sinusoidal function with two different positive amplitudes of around 0.05 and 0.10. Such characteristics in these components attempt to capture the changes in the cycles throughout the observed period, according to the changes detected during the exploratory analysis in Section 6.1. On the other hand, Model 12 has time-dependent parameters including the harmonics in favor of capturing the irregular changes in the amplitudes of the cycle over time. Figure 6.6 Panel (b) shows that the harmonics of Model 12 has changing amplitudes with a large increase from around 1830 to 1850 as well as for around 1890 to 1930. Therefore, the dynamic cycle present in the data should be taken into account and modelled explicitly in order to obtain a more accurate



(a) (b)

Figure 6.6: Harmonics of (a) Model 11 and (b) Model 12

specification of the cyclical behaviour.

Figure 6.7 shows plots of observed versus fitted values for the 12 models. Panels (a) - (d) show the plots for Taylor DBSTAR models 1 - 4, respectively, Panels (e) - (h) for Truncated Power Basis DBSTAR models 5 - 8, respectively, and Panels (i) - (l) for B-splines DBSTAR models 9 - 12, respectively. The closer the data are to the line on the main diagonal of plots, the better the model fits the data. For those models with time-dependent parameters, the fitting performances are the most appropriate. Overall, all 12 DBSTAR models do not show distant points from the lines. Comparing both the standard and periodic DBSTAR models, the latter improve fitting, even though good fittings were observed for the former. It shows, therefore, that DBSTAR models fit Canadian lynx data well.

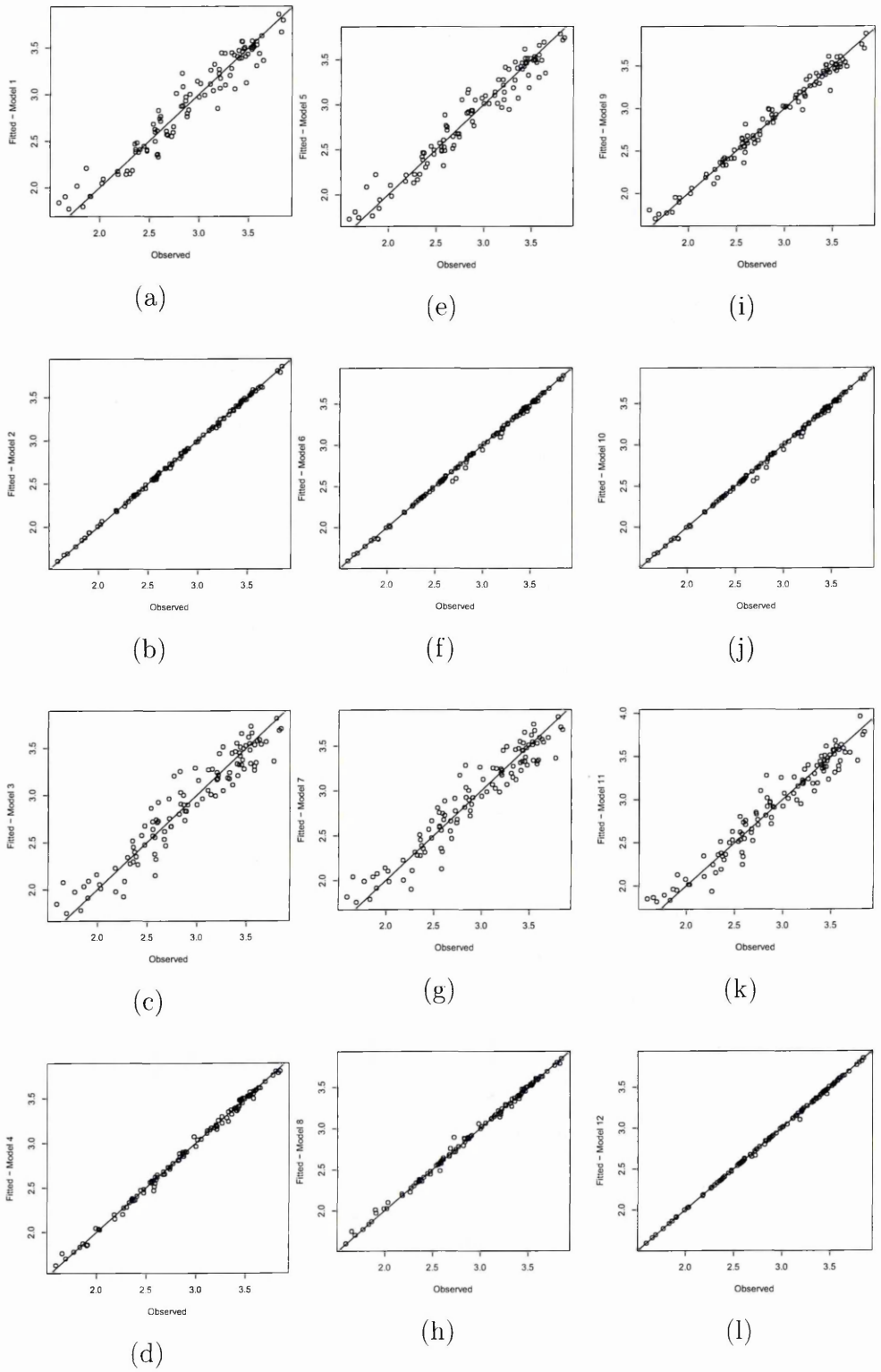


Figure 6.7: Observed versus fitted models - Panels (a) - (d) Taylor DBSTAR models 1, 2, 3 and 4, Panels (e) - (h) Truncated Power Basis DBSTAR models 5, 6, 7 and 8, and Panels (i) - (l) B-splines DBSTAR models 9, 10, 11 and 12

6.4 Comparative model analysis

In this section, 12 DBSTAR models configured in Section 6.3 have their fitting performances compared to the performances of both the classical STAR and CB-STAR models, as in Lopes and Salazar (2005). The CBSTAR uses delay parameter $d = 3$ whilst the classical STAR uses $d = 2$. Both models use autoregressive order $p = 11$. The CBSTAR presents better fitting performance compared to classical STAR models.

Table 6.2 presents the mean absolute errors (MAE) and the root mean squared errors (RMSE) for 12 DBSTAR models as well as for the two competing models, where the measures MAE and RMSE are defined as follows,

$$MAE = \sum_{t=1}^{T} \frac{|e_t|}{T} \tag{6.1}$$

$$RMSE = \sqrt{\sum_{t=1}^{T} \frac{e_t^2}{T}} \tag{6.2}$$

It should be emphasised that static models 1, 3, 5, 7, 9 and 11 were fitted in the interest of making DBSTAR models equivalent to the competitors, since the LSL criterion indicated that the models are dynamic instead. This is because the optimal value of δ_W is less than 1. Out of the 12 models, only models 3 and 7 have large values of MAE and RMSE compared to the CBSTAR model, despite the fact that both models improved fitting compared to the classical STAR model. Both the classical STAR and CBSTAR models have their parameters static over time. On the other hand, all other static model configurations improved fitting. It is worth

Table 6.2: Model comparison - mean absolute errors (MAE) and root mean squared errors (RMSE) - DBSTAR models versus Competitors

Type	Model	MAE	RMSE
Competitors	CBSTAR(11, 3)	0.118	0.153
	Classical STAR(11, 2)	0.142	0.179
Static	1	0.109	0.141
	3	0.135	0.171
	5	0.102	0.134
	7	0.130	0.166
	9	0.061	0.082
	11	0.105	0.139
Dynamic	2	0.012	0.015
	4	0.023	0.032
	6	0.014	0.027
	8	0.019	0.032
	10	0.014	0.027
	12	0.006	0.013

distinguishing B-splines DBSTAR models which showed superior improvements over Taylor series and TPB DBSTAR models. Under other conditions, all the models with time-dependent parameters (2, 4, 6, 8, 10 and 12) present significantly smaller MAE and RMSE compared to the competing models. It shows that the adaptation to the slight changes in pattern from year to year definitely improves fitting performances, regardless the DBSTAR model under analysis.

6.5 Forecasting performance

In this section, an out-of-sample analysis is carried out in order to evaluate the forecasting performances of competing DBSTAR models. The Canadian lynx data set was divided into two parts. The first part is the in-sample period, with observations from 1821 to 1924, while the second part corresponds to the out-of-sample period, which has the last 10 years of the data set, from 1925 to 1934. This division is rather

important to evaluate the forecasting performances of competing DBSTAR models.

When evaluating model forecasting performance, it is essential to take into account not only the performance of point forecasts but also the rigor of their accuracy. Therefore, a measure which determines the accuracy of the joint forecast distribution for the out-of-sample period under investigation, rather than just the point forecasts at each time individually, is preferred. Such a measure is the LPL, defined in equation (3.4). Recall that the larger the value of the LPL, the more evidence in favour for the corresponding model.

Table 6.3 presents the LPL of each DBSTAR model in the out-of-sample period. All the even models show low values of the LPL and evidences for them are little. The LPLs for the even models, i.e., dynamic models, are all lower than the LPLs for the odd models, i.e., static models. This implies that the out-of-sample forecast performance of static models is better than for dynamic models.

Recall that dynamic models allow the parameters to vary in time. During in-sample period, their adaptive behaviour are fast and precise that they present very small forecasting error compared to static models, as shown in previous Section 6.4. However, in the out-of-sample period, their forecasting performance are affected by the parameters' dynamism. Figure 6.8 shows the observed series during out-of-sample period, along with 1-step ahead forecast means and 95% credible interval of Models 3 (a), 7 (b) and 11 (c). The forecast means are, overall, accurate but the forecast uncertainties are overestimated. This overestimation leads to low LPL values. This could be interpreted as an indicative that there may be factors, such as external variables, affecting the variability which are not captured by DBSTAR

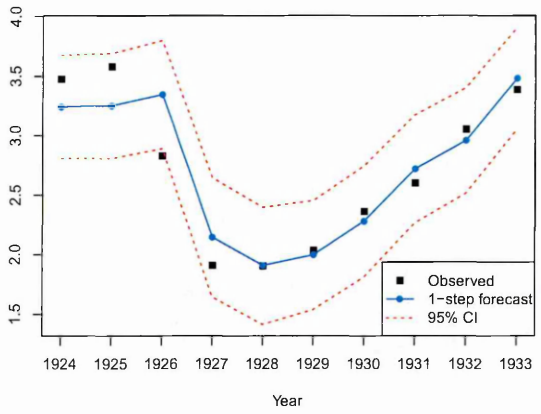
Table 6.3: Out-of-sample log-predictive likelihood (LPL) of each DBSTAR model

Model	1	2	3	4	5	6	7	8	9	10	11	12
LPL	-3.05	-25.40	**0.44**	-12.91	-2.57	-4.43	**0.25**	-11.51	-4.20	-28.00	**-1.89**	-8.09

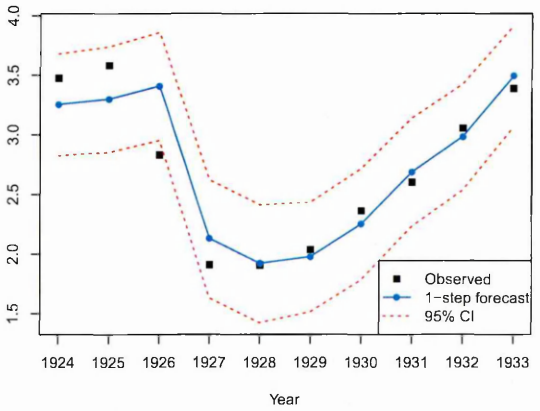
models with time-dependent parameters.

Note that Table 6.3 contrasts the previous analysis from Figure 6.7. On the one hand, fitted models plotted in Figure 6.7 are based on the LSL criterion, which uses filtering algorithm each time an observation becomes available then smoothing algorithm after the last observation is collected to go backward adjusting the estimation of the parameters for the entire sample period. On the other hand, models in Table 6.3 use out-of-sample period only and are based on the LPL criterion. This criterion uses filtering algorithm only, i.e., no backward adjustment is carried out. Therefore, LSL criterion selects models based on fitting performance in order to make them comparable to existing models in the literature (as shown in Table 6.2) whilst LPL criterion selects models based on predicting- performance in order to use them for forecasting purposes.

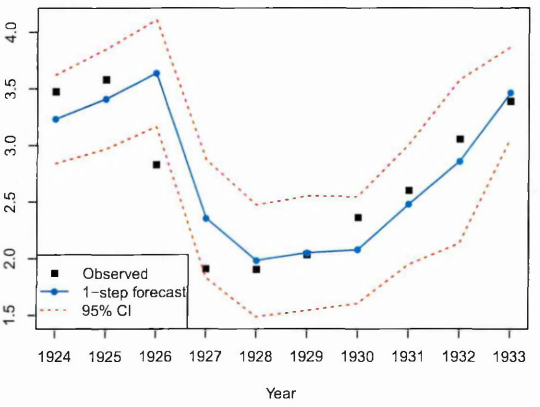
Among all odd models, there are three of them, identified as bold face in Table 6.3, that present the largest values of LPL. They are models 3, 7 and 11, respectively. Recall that these models accommodate periodic components and low autoregressive orders in their structures. Both the observed series in the out-of-sample period, the 1-step ahead forecast means and the 95% credible interval of those three models are illustrated in Figure 6.8. There is periodic behaviour during the out-of-sample period and those three models' intervals cover most of the observation series. Therefore, DBSTAR models are suitable to model the Canadian lynx data set and produce accurate 1-step ahead forecasts.



(a)



(b)



(c)

Figure 6.8: Observed series during out-of-sample period, along with 1-step ahead forecast means and 95% credible interval of Models 3 (a), 7 (b) and 11 (c).

6.6 Discussion

In this chapter, we focused on validating the proposals by capturing the appropriate nonlinear characteristics in order to represent the underlying process. Some of the proposed Taylor and Splines DBSTAR models were applied to the well-known Canadian lynx data set which presents nonlinearity as well as irregular cyclical behaviour. The persistent non-regular cyclic oscillations are apparently asymmetry suggesting the use of nonlinear modelling. Accordingly, different formulations of both the Taylor DBSTAR(r, p) and Splines DBSTAR(r, k, p) models were analysed. The models' parameters were estimated by Kalman filtering as either static or time-dependent. Retrospective analysis using Kalman smoothing was carried out in the interest of making them comparable and the LSL conditional on the parameters was used as a criterion for determination of unknown discrete parameters. Out of 12 DBSTAR models, all of them improved fitting compared to classical STAR models and most of them compared to CBSTAR models.

The analysis during the in-sample period suggests that DBSTAR models with time-dependent parameters should be used to model the Canadian lynx data rather than static DBSTAR models. The former presents adaptive behaviour which is fast and precise, producing then very small residuals and, consequently, smaller MAE when compared to the latter. Since the first two moments of the forecast distributions from the former are updated sequentially over time, its forecasts are more accurate than the latter's forecasts, whose first moment only is updated. Such behaviour is also expected during the out-of-sample period, however the forecast means from both dynamic and static models are, overall, accurate but the forecast

uncertainties from dynamic models seem to be overestimated.

The cycle components changing throughout the observed time series period were appropriately modelled by the time-dependent parameters approaches, including the harmonics in favour of capturing the irregular changes in the amplitude of the cycle in time. We concluded also that standard DBSTAR models attempt to capture the cycle using high autoregressive orders, as do the classical STAR and CBSTAR models, whilst periodic DBSTAR models use low autoregressive orders plus periodic components to model the cycle explicitly.

Compared to the existing approaches, DBSTAR models present a larger number of parameters to be sequentially updated over time. However, the algorithms run in just a few seconds for each model configurations. Hence, the proposals seem promising for real-time applications, as we shall see in the next Chapter 7.

Chapter 7

Forecasting short-term electricity load in Brazil

In this chapter, DBSTAR models are formulated and applied to an hourly time series of electricity load in Southeast and Central-West of Brazil. This is an application that neither classical STAR models nor CBSTAR models are appropriate to be adopted since the underlying series presents nonlinear and non-stationary behaviour including cycles with varying wavelengths and a very large number of data points.

This chapter starts with Section 7.1 describing the electricity sector in Brazil as well as a serious problem that happened in 2001-2002 (and may happen again if the statistical models do not produce accurate electricity load forecasts). The data set used in this application for forecasting analysis purposes is described in Section 7.2.

Section 7.3 reports the exploratory analysis carried out to determine components of the process and the effect of temperature that can help to formulate appropriate DBSTAR models. Section 7.4 specifies three DBSTAR models to be applied to this

data set. In-sample analyses are presented in Section 7.5 and forecasting performances of the models are detailed in Section 7.6. Discussion of this application is reported in Section 7.7.

7.1 The Brazilian electricity market

The electricity market in Brazil is highly dependent on hydroelectricity generation, which meets over 80% of its electricity demand. Brazil has the largest capacity for water storage in the world, the largest electricity market in South America, and a large transmission network across regions (Soares and Souza, 2006).

This dependence on hydropower makes the country exposed to power source deficit in dry periods. Controlling the electricity market is a task rather challenging not only in Brazil but worldwide as the electricity system operators need to balance the power production and the demand. Therefore, it is crucial for the systems to optimise the processes by setting up a program for generation and transmission, including hourly demand load forecasts for various time horizons. Thus, the more accurate electricity load demand forecast the better, or less risky, are the contracts for both the generators and the distributors. Among the most important time horizons for forecasting hourly loads are those from 1 hour to 168 hours (one week's time) ahead.

The National Interconnected System (in Portuguese, Sistema Interconectado Nacional, SIN) constitutes the electricity organisations in the North, Northeast, South, Southeast and Central-West regions of Brazil. The electricity is freely transported within each region. In this thesis, we focus on the two most important regions,

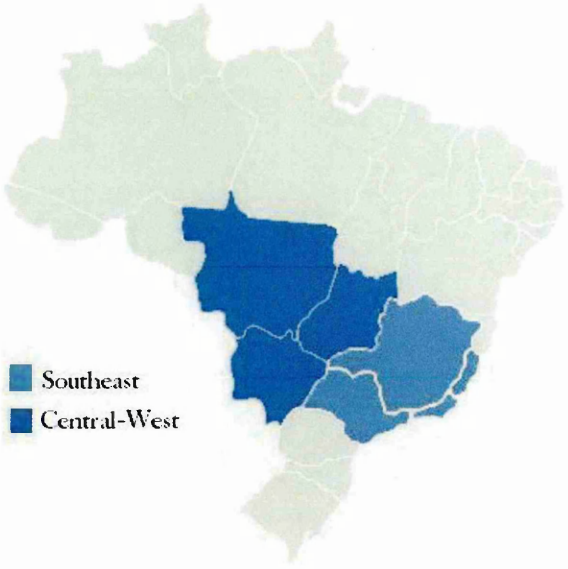


Figure 7.1: Brazil map - Southeast and Central-West regions

Southeast and Central-West. The total population in both regions is 94,422,504 representing 49.5% of the total Brazilian population. Figure 7.1 illustrates the two regions under analysis.

7.1.1 Electricity crisis in 2001-2002

Brazil had a serious electricity generation and transmission crisis in 2001-2002. The country had an electricity rationing period that lasted between June 2001 and February 2002, having impacts in several sectors in its economy.

Before the electricity generation and transmission crisis in 2001-2002, consumption had been found to be increasing at similar rates to the Brazilian economy, whereas the installed capacity (the capacity to deliver power at a given time) had been also increasing, but at a slower rate. Brazil had already planned to have a privatisation model of the electricity sector (Perobelli and de Oliveira, 2013) with

the creation of a national electricity regulatory agency, Agência Nacional de Energia Elétrica (ANEEL), with a national system controller, Operador Nacional do Sistema (ONS). For more details on the Brazilian electrical system reforms, see Mendonça and Dahl (1999).

According to Rosa and Lomardo (2004), before these reforms implemented prior to the crisis, Brazil would not have supported any expansion to generate more electricity. Blackouts had already occurred a few times, indicating the nearing crisis. The level of water stored in reservoirs were lower year after year and supply could no longer be ensured. From 1995 to 2000, the maximum capacity of the reservoirs reached an average of at least 96%, 89%, 77%, 88%, 83%, 70% and 59%, successively, of its capacity each year (Rosa and Lomardo, 2004).

In 2001, the Brazilian Government admitted the electricity crisis. This was caused not only by the scarcity of rain, but also a lack of investment needed to expand generation and transmission capacity. A compulsory electricity rationing program was implemented for eight months in order to avoid more blackouts. So, between June 2001 and February 2002, consumers were encouraged to save electricity by being given bonuses rewarded for consumption well below the target or being penalised for over-consumption otherwise. After that period, the country successfully achieved the goal of reducing the total consumption level.

After the electricity crisis, the reforms (implemented prior to the crisis) were tightly controlled with greater rigidity by ONS. Centralisation and regulation of supply, based on an electricity auction system, were the main instruments to guarantee future supply to consumers (Mendonça and Dahl, 1999). The implemented

auction system increased efficiency and attracted private capital investment to the electricity sector and created futures market for long-term electricity supply. Contracts for future supply introduced competition in the power generation segment for better agreements between generators and distributors (Moita, 2008). It constituted fair, solid and transparent regulations, planned to provide supply and the sustained extensions of the main intrinsic power sector activities - generation, transmission and distribution.

7.1.2 The need for electricity load forecasting

Within the scenario of the Brazilian electricity sector privatisation, it became essential for providers to plan their purchases in advance. Electricity supply auctions are contracted for three to five years ahead of provision. The regulatory agency ANEEL required the providers to inform them of their load requirements and contract the required load entirely, subject to penalties in case of large forecast deviations. Problems of balance between electricity demand forecasts and production had been frequent before the reforms.

There are different types of electricity load forecasting methods categorised according to the forecast horizons. Short-term electricity load forecasting, usually observed hourly or half-hourly, are based on high frequency data for producing forecasts for the next (half-) hour up to the next week or even 10 days.

Longer electricity load forecasting horizons are predominantly used for economic modelling. They help making decisions for maintenance programs investment planning, which helps the generation process to obtain optimal production and also the

distributors to extend their networks. Generally, the mid-term forecasting horizons are between a month and a year and the long-term forecasting horizons are longer than 1 year.

Producing accurate electricity load forecasts is fundamental to guaranteeing the enduring balance between electricity demand and production in the network, and help to avoid big problems such as the electricity crisis in 2001-2002.

7.2 Electricity load data set description

The electricity data set analysed in this thesis consists of the hourly electricity load, measured in MegaWatts (MW), and temperature, in degree Celsius ($°C$), in the Southeast and Central-West regions of Brazil from the first hour on 1 June 2003 to the last hour on 30 June 2010, giving therefore a total of 7 years of data, or 62,088 hourly observations. The electricity load data are aggregated while temperature data are averaged across the states in both Southeast and Central-West regions of Brazil.

In Brazil, both Southeast and Central-West regions change time for the Brazilian Summer Time (BRST) around October/November and change it back to standard Brazilian Time (BRT) in February. The data set registers 23 hourly observations for the days that BRST starts (19 Oct 2003, 2 Nov 2004, 16 Oct 2005, 5 Nov 2006, 14 Oct 2007, 19 Oct 2008 and 18 Oct 2009), when the BRT is about to reach midnight, clocks are turned forward 1 hour to 1AM local BRST. Similarly, there are 25 hourly observations for the days it ends (15 Feb 2004, 20 Feb 2005, 19 Feb 2006, 25 Feb 2007, 17 Feb 2008, 15 Feb 2009 and 15 Feb 2010), as when the BRST is about to

reach midnight, clocks are turned backward 1 hour to 11PM BRT instead.

With a view to avoiding complications such as missing data or duplication of data for the same hour in the data set, this changing time is not taken into consideration. In such a way, the data set is formed with 24 hourly observations throughout the period under investigation.

We also obtained external variables such as one to indicate the days of the week and another for holidays and bridge-holidays (defined below). In Brazil, there are National holidays, such as Christmas Day, New Year's Day, Independence Day, Labor Day, Children's Day, and so forth, and local holidays. We consider in this thesis only National holidays since the data are aggregated from two large regions of Brazil. Notice that a bridge-holiday, generally a Monday or a Friday, is the day before a holiday preceded by a weekend or the day after a holiday followed by a weekend. For instance, if the holiday is on a Tuesday, the Monday immediately before is considered a bridge-holiday. Or if the holiday is on a Thursday, the Friday immediately after is considered a bridge-holiday.

The data set under analysis was divided into two groups, denoted henceforth in-sample and out-of-sample periods. Table 7.1 illustrates the sample divisions. The first group starts the observation period at the first hour on 1 June 2003 to the last hour on 31 May 2010 and is referred to as the in-sample period. Analysis in this period is reported in Section 7.5. The second group is the observation of the whole last month of the data set, i.e., from the first hour on 1 June 2010 to the last hour on 30 June 2010 and is referred to as the out-of-sample period. The reason this period was reserved as out-of-sample is to evaluate the forecasting performances of

Table 7.1: Sample division

Period	Start	End	Observations
In-sample	1 June 2003	31 May 2010	61,368
Out-of-sample	1 June 2010	30 June 2010	720
Total	1 June 2003	30 June 2010	62,088

competing DBSTAR models, as we shall see in Section 7.6. Both in-sample and out-of-sample periods include all characteristics present in this type of data set, to challenge DBSTAR models to perform well in both periods.

7.3 Exploratory data analysis

This section describes the generic patterns of the electricity load series to determine components of interest to be considered for inclusion in the structure of DBSTAR models for the underlying process. Each component should have an explicit stochastic formulation and a direct interpretation.

Figure 7.2 shows a plot of the average hourly electricity load (MW) by days of the week plus holidays (Hol) and bridge-holidays (BH) in the Southeast and Central-West regions of Brazil. Three distinct groups of intra-day patterns can be observed: weekdays, Saturdays plus holidays and Sundays plus bridge-holidays. On average, the consumption of electricity has similar patterns and levels from Monday to Friday, with slightly lower values on Mondays. There are different patterns for holidays and bridge-holidays. Recall that a bridge-holiday is always a weekday.

In general, the consumption of electricity decreases during the first 5 hours of the day regardless of the day of the week. On Sundays and bridge-holidays, the consumption keeps low, reaching lowest values until around 8-9AM. From Monday

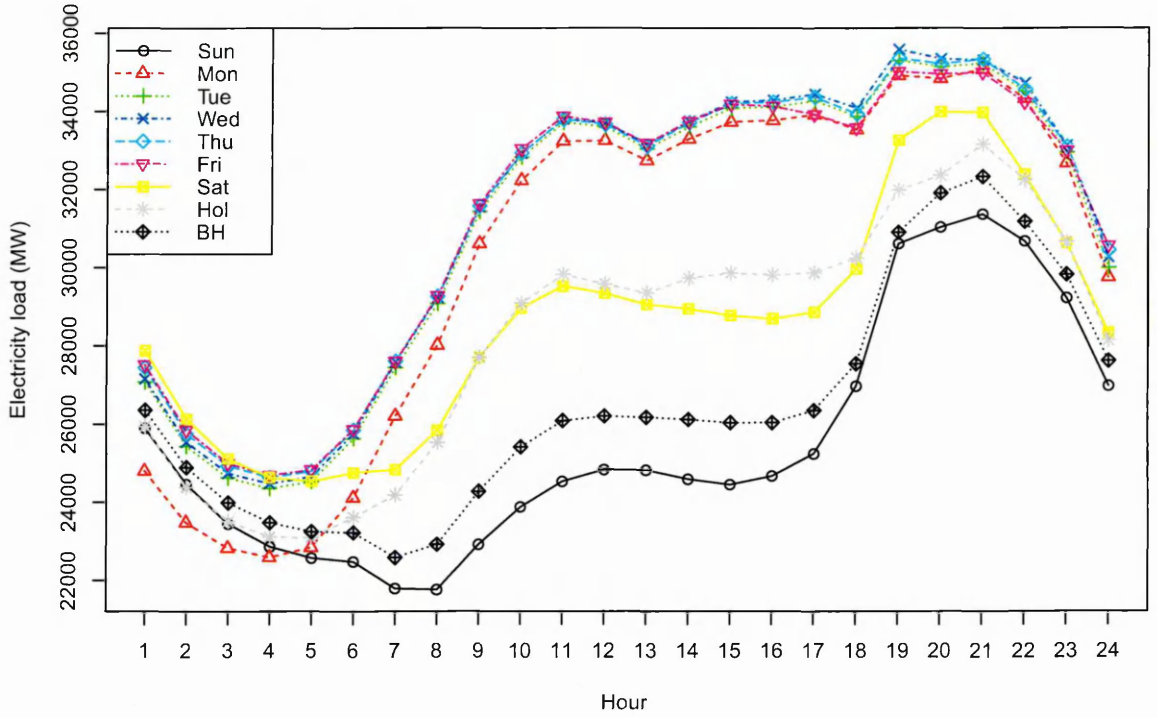
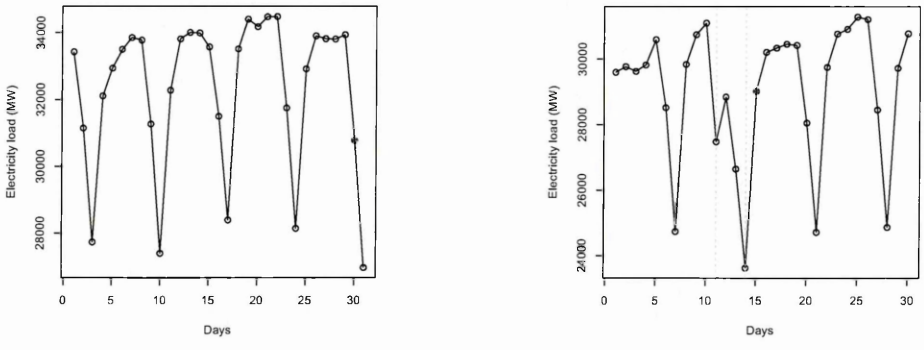


Figure 7.2: Average hourly electricity load (MW) by days of the week plus holidays and bridge-holidays - Southeast and Central-West regions

to Friday, due to normal activities such as work, school, university, the load starts to increase earlier than on Sundays and bridge-holidays, when most of the consumers starts their routines slightly later. The consumption also decreases at the end of the day, from around 9PM. Also, the demand for electricity appears to remain at same levels around lunch time for all days of the week. After lunch, the consumption increases back and remains high until 8-9PM. Peak times occur from 6PM to 9PM.

Holidays

Due to the often idiosyncratic behaviour of holidays and bridge-holidays, generally the forecasting error on those special days are significantly higher than on normal



(a) (b)

Figure 7.3: Average daily electricity load in (a) June 2008 and (b) June 2009. Notice that the shape of the curve (b) changes due to holiday and bridge-holiday, as identified between the two dashed lines

weekdays (Kim, 2013). Figure 7.3 illustrates a case when those changes are very noticeable. Panel (a) displays the average daily electricity load in June 2008 and panel (b) in June 2009. Notice that the shape of the curve in panel (b) changes due to a holiday and a bridge-holiday, as shown between the two dashed lines, which identify the electricity load from Thursday to Sunday. On Thursday 11 June 2009, it was the *Corpus Christi* (Latin for Body of Christ) holiday, a Catholic celebration on the Thursday 60 days after Easter. It is a mobile holiday as Easter varies from year to year. Notice that there is a drop in load on that day. A similar change was detected on the day after, Friday 12 June 2009, as it was a bridge-holiday. Comparing both days with the corresponding days the year before, Thursday 12 June 2008 and Friday 13 June 2008, we detect that their electricity load were typical for Thursdays and Fridays. Therefore, the inclusion of holiday information in the modelling stage, using a predictor dummy variable to identify holidays and bridge-holidays, should be considered.

Nonlinearity caused by temperature

Weather conditions play an essential role on the behaviour of electricity load, especially temperature. Generally, a nonlinear association between them is detected, as illustrated in Figure 7.4, which shows the average weekly electricity load (MW) versus temperature in degree Celsius ($°C$) in the Southeast and Central-West regions of Brazil. Notice that to better analyse such behaviour, the data were aggregated to a weekly basis, since the number of observations collected hourly is very large.

The electricity load increases when temperature is increased but in a nonlinear fashion. Typically, in countries where very low temperatures are recorded during Winter time, a U-shape is observed in the scatterplot of electricity load versus temperature (see, Pardo *et al.* (2002) and Dordonnat *et al.* (2008)). In those countries, the electricity load is typically high for very cold and very hot days. On the other hand, although very cold temperatures were not registered in Southeast and Central-West of Brazil in the period under analysis, the temperature effect on electricity load in those regions is not constant across time either. The minimum and maximum registered temperatures were 11 $°C$ and 39 $°C$, respectively, and for cold and mild temperatures, the load increases slower than for high temperatures. Notice that, the influence of temperature on load is much larger with warmer temperatures (above 25 $°C$). This analysis indicates that an S-shape pattern is present rather than a U-shape for the Brazilian data.

In addition, temperature seems to be more closely related to the electricity load yearly cycle, better than information coming from a categorical variable, such as seasons of the year, for very disaggregated data (i.e., hourly observations). Therefore,

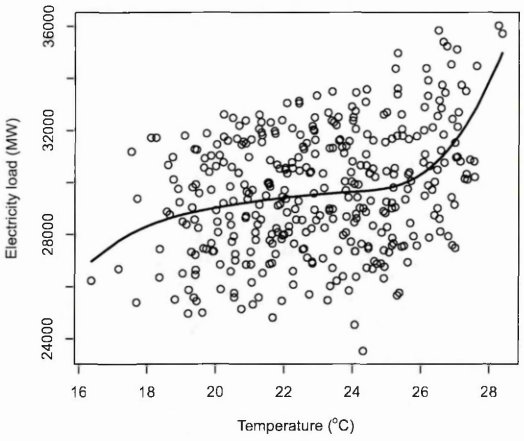


Figure 7.4: Average weekly electricity load (MW) versus temperature in degree Celsius ($°C$) - Southeast and Central-West regions. The curve indicates a S-shape pattern rather than a U-shape for the Brazilian data, which suggests a nonlinear relationship between electricity load and temperature.

in this analysis, temperature is used as the transition variable (s_t) in the modelling stage, due to its nonlinear association with electricity load (Y_t).

Non-stationarity

Although hourly data show a rather weak either upward or downward trajectory, a more pronounced increasing pattern was detected in a more aggregated data, such as average monthly electricity load, as shown in Figure 7.5. It shows that electricity load is a long-term non-stationary process. Ideally, an effective model would require a trend component in its structure. Long-term non-stationarity may be modelled by DBSTAR models with the inclusion of such a component.

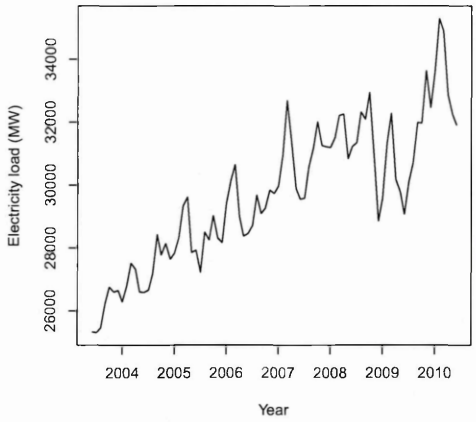


Figure 7.5: Average monthly electricity load in MegaWatts (MW) - Southeast and Central-West regions

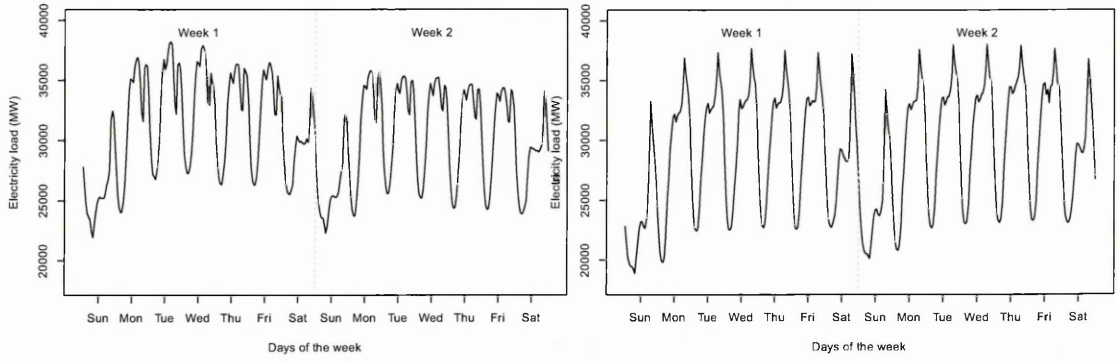
Cyclical patterns

Figure 7.6 reveals typical hourly electricity load curve shapes according to the season of the year. Panel (a) illustrates a fortnight in January 2009, therefore summer, and panel (b) a fortnight in June 2009, winter time.

Notice that the shape of the electricity load curve at peak times varies according to the season. During winter, there is a sharp aspect in the curve in the evenings which is not present in the summer curve due to the BRST that encourages the use of natural lights early in the evenings.

A within-day cyclical pattern of wavelength 24 hours is detectable from the similarity of the electricity load from one day to the next, particularly on weekdays. Similarly, a within-week cyclical pattern of wavelength 168 hours is visible when the electricity load of one week is compared to the corresponding weekday of neighbouring weeks.

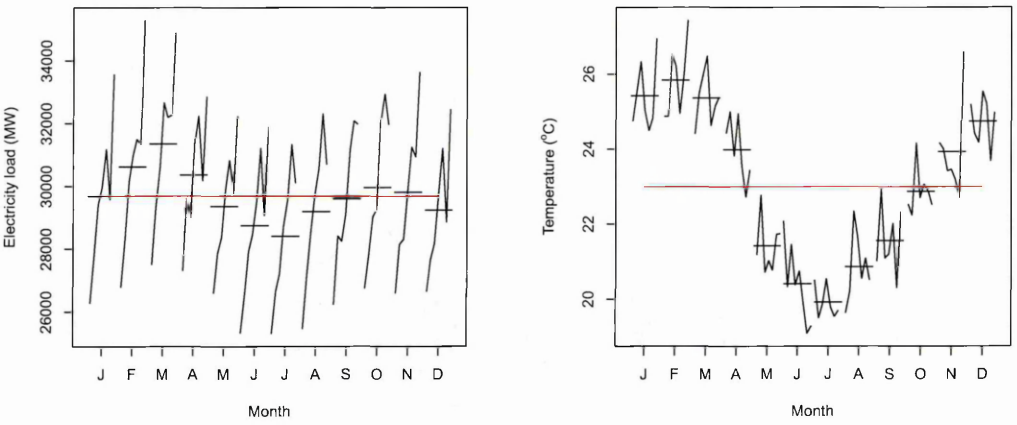
It is also noticeable that a cyclical pattern exists from one year to the next,



(a) (b)

Figure 7.6: Typical hourly electricity load curve shapes according to the season of the year, summer and winter, respectively. (a) from Sunday, 11 January 2009 to Saturday, 24 January 2009 and (b) from Sunday, 14 June 2009 to Saturday, 27 June 2009.

as illustrated with average monthly electricity load and temperature in Figure 7.7. Panel (a) shows a monthplot of average monthly electricity load in MegaWatts (MW), and panel (b) shows a monthplot of temperature in degree Celsius ($^{\circ}C$) - Southeast and Central-West regions. How to read the monthplot? The monthplot allows us to detect both between-month and within-month cyclical patterns. The months are displayed on the X axis and the response time series on Y axis. The long red horizontal line is the average for the entire period of 7 years. In practice, this red line should only be used as a reference line. The 12 short black horizontal lines are averages of each 12 months of the sample period. The within-month curves are the monthly average for each of the 7 years. The figure shows that electricity load and temperature hourly data present cyclical patterns with different shapes. For temperature, it is quite clear that it follows a U-shape aspect, being high during summer and low during winter, and with transition behaviours during the other two seasons, spring and autumn. We detected that for the electricity load, however,



(a) (b)

Figure 7.7: Panel (a) Month plot of average monthly electricity load in MegaWatts (MW) and panel (b) Month plot of temperature in Degree Celsius ($°C$) - Southeast and Central-West regions. The red lines are the average for the entire year.

a S-shape pattern is present rather than a U-shape due to the BRST with the use of natural lights early in the evenings. Observe that the monthly electricity load means are close to the mean for the whole year. Also, during the BRST (October, November, December, January and February), the loads are relatively low. In addition, these cyclical pattern changes from year to year were detected, so we consider it to be a non-stationary behaviour. This distorted pattern should also be taken into account for modelling (via transition variable temperature), which we expect to represent the yearly cycle more accurately than seasonality itself.

Hence, cycles with different wavelengths should be taken into account in order to model the cyclical patterns present in this data set in an appropriate way. To analyse each of the cyclical effects, three substances are considered: for the daily cycle, harmonic components of wavelength 24 periods, for the weekly cycle, harmonics components of wavelength 168 periods and finally for the yearly cycle, temperature is used as the transition function in DBSTAR models. As it was observed that

the three electricity load patterns change with time, particularly in the periods of changing seasons, the harmonics in the models are time-dependent components.

7.4 DBSTAR model formulation

Following the developments reported in the previous Section 7.3, let $Y_t, t = 1, \ldots, T$, be the electricity load time series and s_t be the temperature at each hour for the whole sample period. DBSTAR models should use the following modelling structure:

$$Y_t = NonlinearAR(s_t) + Trend + Cycle_{day} + Cycle_{week} + Holidays + \xi_t. \quad (7.1)$$

Nonlinear $AR(s_t)$ represents the polynomial part of the model, either the Taylor series defined in Section 4.2 or Splines defined in Section 4.3, with the use of temperature as the transition variable. *Trend* is the component for non-stationarity, as defined in Section 5.1. $Cycle_{day}$ represents the within-day cycle, $Cycle_{week}$ the within-week cycle, and both components are incorporated into the models as defined in Section 5.2. *Holidays* is the dummy covariate for holidays and bridge-holidays which is incorporated in DBSTAR models as described in Section 5.3, and finally, $\xi_t \sim N(0, \sigma_t^2)$ is the error term of the model, independent and identically normally distributed.

In this section, three suitable DBSTAR formulations are applied to the electricity load data set, denoted henceforth by Model A, Model B and Model C. Model A uses Taylor series to approximate the nonlinear transition function, as defined in Section 4.2, Model B uses Truncated Power Basis (TPB) and Model C uses B-

splines function, both defined in Section 4.3. They are set similarly to those models for the Canadian lynx time series application, which were described in Section 6.2. Nonetheless, there are some particularities in the application to the electricity data set which are distinguished in this section. The mathematical formulation of the modelling structure in (7.1) is given as,

$$\left(Y_t \mid \underline{\theta}_t, \underline{\tau}_t, \underline{\psi}_{1t}, \underline{\psi}_{2t}, \beta_t\right) \sim N\left(\underline{F}'_{1t}\underline{\theta}_t + \underline{F}'_{2t}\underline{\tau}_t + \underline{F}'_{3t}\underline{\psi}_{1t} + \underline{F}'_{4t}\underline{\psi}_{2t} + F_{5t}\beta_t, V_t\right), \quad (7.2)$$

$$\left(\underline{\theta}_t \mid \underline{\theta}_{t-1}\right) \sim T_{n_{t-1}}\left(\mathbf{G}_{1t}\underline{\theta}_{t-1}, \mathbf{W}_{1t}\right), \quad (7.3)$$

$$\left(\underline{\tau}_t \mid \underline{\tau}_{t-1}\right) \sim T_{n_{t-1}}\left(\mathbf{G}_{2t}\underline{\tau}_{t-1}, \mathbf{W}_{2t}\right), \quad (7.4)$$

$$\left(\underline{\psi}_{1t} \mid \underline{\psi}_{1t-1}\right) \sim T_{n_{t-1}}\left(\mathbf{G}_{3t}\underline{\psi}_{1t-1}, \mathbf{W}_{3t}\right), \quad (7.5)$$

$$\left(\underline{\psi}_{2t} \mid \underline{\psi}_{2t-1}\right) \sim T_{n_{t-1}}\left(\mathbf{G}_{4t}\underline{\psi}_{2t-1}, \mathbf{W}_{4t}\right), \quad (7.6)$$

$$\left(\beta_t \mid \beta_{t-1}\right) \sim T_{n_{t-1}}\left(G_{5t}\beta_{t-1}, W_{5t}\right). \quad (7.7)$$

The distribution (7.2) is given for each hourly electricity load (Y_t), with the state vector $\underline{\theta}_t$ to accommodate each polynomial coefficient associated with the polynomial regression variables in $\underline{F}_{1t}$, following the multivariate Student-T distribution specified in (7.3). The trend component uses a two-dimensional vector $\underline{\tau}_t$ to detect non-stationarity and is referred to as local level plus local linear growth rate, following a multivariate Student-T distribution as in (7.4). The components for modelling the cycles are linear combination of trigonometric terms for representing the daily ($\underline{\psi}_{1t}$) - wavelength 24 hours - and weekly ($\underline{\psi}_{2t}$) - wavelength 168 hours - periodic

behaviour and are distributed as in (7.5) and (7.6), respectively. Finally, the coefficient β_t is associated with the dummy covariate for holidays and bridge-holidays, following a Student-T distribution as in (7.7).

All the parameters in the model (7.2) are updated sequentially in time, as shown in Section 3.2.2, including the unknown observational variance V_t of the underlying nonlinear AR process as previously detected to be stochastically changing, as well as each state covariance matrix $\mathbf{W}_{it}, i = 1, \ldots, 4$ and the scalar state variance W_{5t}. For this reason, discount factors δ_V for the observational variance V_t and δ_W for the state covariance matrices $\mathbf{W}_t$, where $0 < \delta_V \leq 1$ and $0 < \delta_W \leq 1$, are used in this analysis.

Non-informative prior distributions were also used to form initial priors for all model parameters at time $t = 0$, as we did not have any prior expert knowledge about them. Thus, we set $\underline{m}_0 = \underline{0}$ and $\mathbf{C}_0 = 3\mathbf{I}$, with $\mathbf{I}$ the Identity matrix, as the initial prior mean vector and covariance matrix, respectively.

The polynomial regression variables from $\underline{F}_{1t}$ vary according to the models specifications, as follows:

- Model A (Taylor series): $\underline{F}'_{1t} = [\underline{z}'_t, \underline{z}'_t s_t, \underline{z}'_t s_t^2, \underline{z}'_t s_t^3]$;
- Model B (TPB): $\underline{F}'_{1t} = [\underline{z}'_t, \underline{z}'_t s_t, \underline{z}'_t s_t^2, \underline{z}'_t s_t^3, z_t(s_t - \kappa_1)_+^3, \ldots, z_t(s_t - \kappa_k)_+^3]$;
- Model C (B-splines): $\underline{F}'_{1t} = [\underline{z}'_t, \underline{z}'_t B_1(t), \ldots, \underline{z}'_t B_m(t)]$.

The autoregressive degree p, within the vector $\underline{z}_t$ is selected using the Bayesian approach based on the predictive performance of different models, described in Section 3.3 and should not be high, since the models have components for all cyclical patterns previously mentioned.

7.5 In-sample analysis

In this section, the results of the implementation of model (7.2) are reported. The parameter estimation is based on observations from the first hour on 1 June 2003 to the last hour on 31 May 2010, therefore 7 years of hourly data.

Kalman filtering was implemented using the software R (version 2.15.2) and run using a Linux server. The algorithms had to be run using the server because the software R on a PC Desktop (Intel Core 2 Duo CPU, 2.93GHz, 2.96GB of RAM, Windows XP) could not allocate vectors or matrices of high dimensions. The dimensions are large in this application as it is a high frequency data set. It takes approximately 2 minutes to run each algorithm, hence, the proposed models are suitable for real-time applications.

The idea is that rather than a least squares approach or an MCMC approach, Kalman filter algorithm is in use for estimating the parameters of the dynamic models based on recorded observations. This technique considers the accuracy of the predictions computed for the observations at each time, which explicitly exploits the dynamical structure of the studied models.

7.5.1 Configuring DBSTAR models

The log-predictive likelihood (LPL) was used as a criterion for determination of unknown discrete parameters of DBSTAR models. In this application, the aim of the DBSTAR model is prediction, therefore it makes sense to determine the model based on prediction error rather than model fit such that the prediction error would be minimal. Several models were run which differed only in values of parameters

that were not accommodated into the state vectors, such as, autoregressive order p and delay parameter d, and those were selected by the LPL conditional on the parameters based on models' forecasting performance, as defined in Section 3.3.2. The LPL criterion gives evidence in favour of a model which presents the largest value of LPL.

Table 7.2 presents the final DBSTAR model configurations. For approximating the logistic function in (2.13) as the smooth transition function, the Taylor series degree is chosen $r = 3$ and both the TPB and the B-splines use polynomial degree $r = 3$. The AR order is selected as $p = 1$, as the models have components for all cyclical patterns previously mentioned, so no high AR order was expected. Temperature was used as the transition variable (s_t) and the delay parameter d for (s_{t-d}) is selected as $d = 1$ for all model configurations. The delay of an hour corresponds to the time that consumers react to the change in temperature to consume more (or less) electricity.

The LPL criterion selects only 1 knot for both the TPB and the B-splines, which is located at the median temperature 23^oC. The LPL criterion also selects both $\delta_W = \delta_V = 0.99$. Notice that the parameters of these models vary in time, but slowly as the discount factor is very close to 1. This means that from time $t-1$ to time t, both the observational and the state variances get discounted by 1%, i.e., the model only brings 99% of the variability from the immediate time before (static models would bring 100%). In addition, the LPL selected two harmonics for each of the daily and weekly cyclical patterns to represent the periodic patterns in the series.

Table 7.2: Final DBSTAR model configurations for the electricity load data set

	Model A	Model B	Model C
Polynomial degree	3	3	3
AR order	1	1	1
Delay	1	1	1
Harmonics (day)	2	2	2
Harmonics (week)	2	2	2
Knots	-	1	1
δ_W	0.99	0.99	0.99
δ_V	0.99	0.99	0.99
No. param.	21	27	23

Models A, B and C present very similar configurations and the results obtained individually are also very similar. Model C was chosen to be analysed as it presented the largest LPL during the out-of-sample period.

7.5.2 Modelling the components of DBSTAR models

Observational standard deviation estimate

The average (by days of the week plus holidays and bridge-holidays) of the estimated observational standard deviations ($\sqrt{S_t}$) from equation (7.2) are presented in Figure 7.8. Recall that a discount factor δ_V for the observational variance was specified so the estimated variance is allowed to change sequentially in time.

Generally, the standard deviations present very similar patterns to the hourly electricity load, shown in Figure 7.2. On average, the variability of electricity load has similar patterns from Monday to Friday, decreasing successively, contrasting with the shapes of the curves at weekends and bridge-holidays. At weekends and bridge-holidays, there is a sharp increase in the variability at peak times, raising the forecast uncertainty. The higher the standard deviations the wider the credible

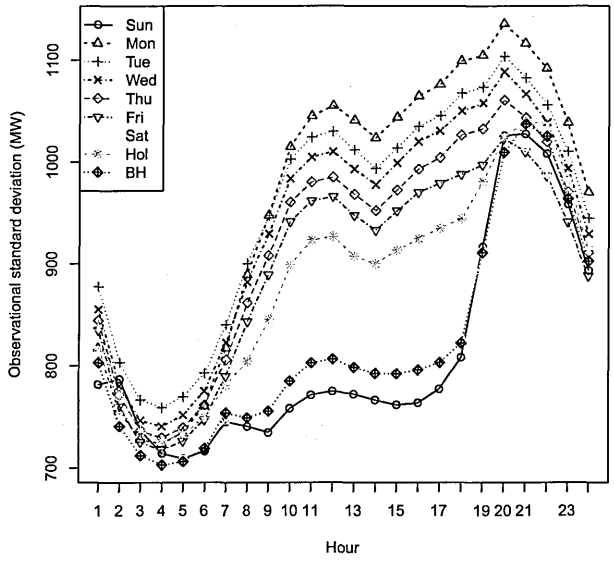


Figure 7.8: Average in-sample hourly observational standard deviation estimate ($\sqrt{S_t}$) in MW by days of the week plus holidays and bridge-holidays.

intervals, as we will see in Section 7.6. Overall, the electricity load cycles vary more for weekdays than for weekends and holidays.

Regardless of the day, the electricity load presents higher variability at peak times (from 6PM to 9PM), consequently, it is more difficult for the models to predict the load in that period, when the credible interval of the forecasts is wider, than during off peak hours.

Posterior mean of AR coefficients ($\underline{\theta}_t$)

Figure 7.9 illustrates the posterior mean of coefficients $\underline{\theta}_t$ associated with the polynomial regression variables. A discount factor δ_W for each of the state vectors was specified so the estimated coefficients evolve dynamically in time.

The first coefficient, $\underline{\theta}_{1t}$, represents the intercept and local non-stationarity can

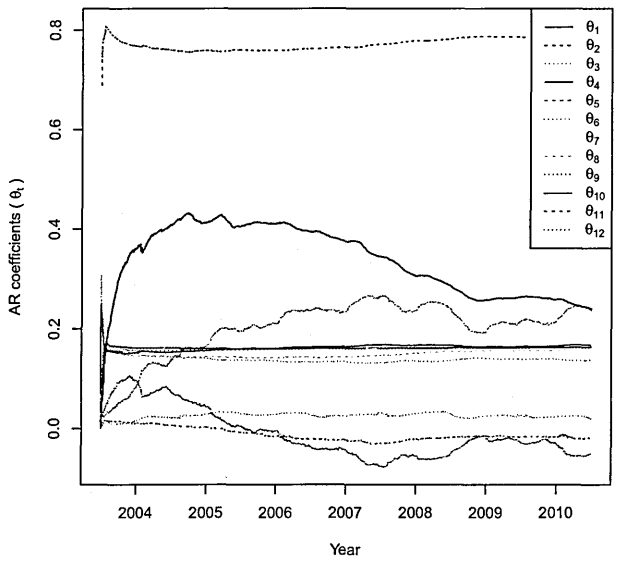


Figure 7.9: In-sample posterior mean of coefficients associated with polynomial regression variables.

be represented by this coefficient. The second coefficient, $\underline{\theta}_{2t}$, is associated with a lagged value of electricity load (y_{t-1}). This coefficient can be interpreted as the contribution rate to the electricity load from the immediate load before. The effects of both $\underline{\theta}_{1t}$ and $\underline{\theta}_{2t}$ are positive and more pronounced for the whole sample period compared to the other coefficients.

The remaining coefficients $\underline{\theta}_{it}, i = 3, \ldots, 12$, take into account the nonlinearity present in the data as they are associated with the interaction between electricity load and temperature. Notice that the posterior means of most of them are significantly different from zero, oscillating around 0.10 and 0.20, which suggests that the nonlinearity is represented by the model. The transition variable temperature also presents a yearly cycle, which confirms the nonlinear and periodic nature of these time-dependent coefficients.

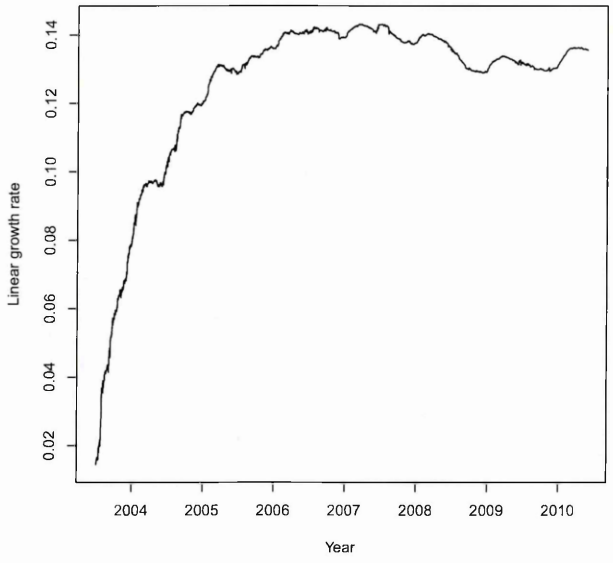
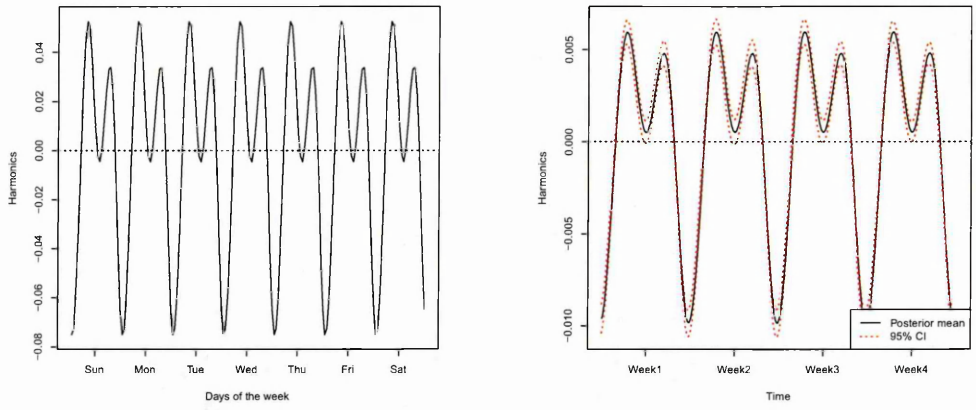


Figure 7.10: In-sample posterior mean of the trend component - the linear growth rate.

Posterior mean of the trend component ($\underline{\tau}_t$)

Figure 7.10 presents the time variation in the posterior mean of the trend component of the vector $\underline{\tau}_t$. One of the elements in this vector can be interpreted as the linear growth rate. The posterior mean of the linear growth rate parameter rises sharply from around 0.01 up to 0.14 between 2004 and 2006, and remains at around same level until the end of the period. Global non-stationarity is well captured by the model, in accordance with the exploratory analysis pointed out in Figure 7.5. The estimated values are always positive which indicates an increase in the electricity load for the whole period. This analysis ensures that both local and global non-stationarities, the former is modelled by time-varying AR components whilst the latter is modelled by the trend component, are appropriately modelled by the DBSTAR approach with positive growth at non-constant rates.



(a) (b)

Figure 7.11: In-sample posterior mean of the periodic components to represent (a) daily cycle and (b) weekly cycle with 95% credible interval. The period in (a) is from the first hour on 11 Jan 2009 to the last hour on 17 Jan 2009, and in (b) is from the first hour on 04 Jan 2009 to the last hour on 31 Jan 2009.

Posterior mean of the periodic components $(\underline{\psi}_{1t})$ and $(\underline{\psi}_{2t})$

Figure 7.11 shows the in-sample posterior means of the periodic components: Panel (a) shows the daily cycle and Panel (b) shows the weekly cycle, with 95% credible intervals. The reason the daily cycle is not plotted with 95% credible intervals is that the intervals are quite tight. And the reason they are quite tight is because the periodic repetitions within each day are very well behaved. Nevertheless, this periodic component is statistically significant and should be kept in the model structure. The credible interval for the weekly cycle can also be considered narrow, but the presence of holidays and bridge-holidays increases some variability.

From Panel (a), negative effects are obtained for the first hours of the day and late nights. Those are the periods when the consumption of electricity are minimal. Positive effects are detected in the mornings and at peak times, with a drop in the afternoons. Such behaviour was identified during exploratory analysis as shown in

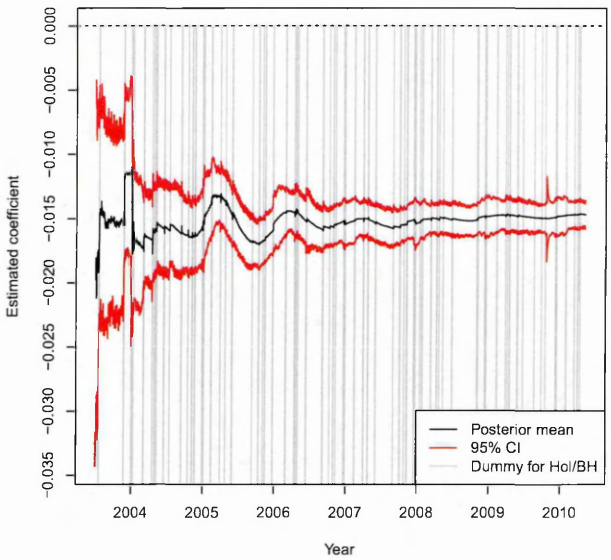


Figure 7.12: In-sample posterior mean of the component for holidays and bridge-holidays.

Figure 7.6. In Panel (b), negative effects are obtained for weekends and positive effects for weekdays, also identified in Figure 7.6 during exploratory analysis. These analyses indicate that DBSTAR models are appropriately modelling these patterns.

Posterior mean of the coefficient (β_t) for holidays and bridge-holidays

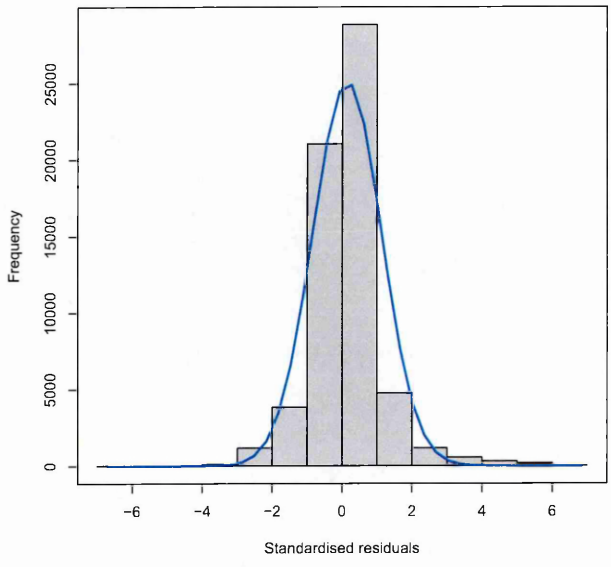
Figure 7.12 shows the in-sample posterior means of the component for holidays and bridge-holidays together with 95% credible intervals as well as the dummy covariate (vertical grey lines when there is a holiday or bridge-holiday). This statistically significant coefficient affects the electricity load negatively, as the posterior mean is below zero for the entire period. It is in agreement with the exploratory analysis, as the electricity load shows a decrease during holidays and bridge-holidays, as shown

in Figures 7.2 and 7.3. Accordingly, the effect of this coefficient was expected to be negative. Notice that both the posterior means and variances get updated only at the times identified as vertical grey lines. The high variability of this component at the start is because the model is learning and only being updated on a few holidays in that period. The holiday effects are essential in this application and are also notably at around the same level over time, especially from 2006 on.

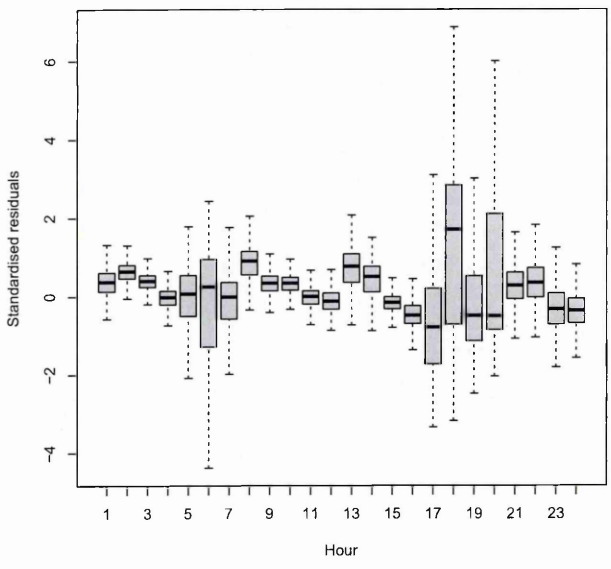
7.5.3 In-sample 1-step ahead forecast error diagnostic

Various diagnostic tools can be used to investigate the residuals. Figure 7.13 shows a histogram together with a normal density function (blue curve) in Panel (a) and a boxplot for each hour in Panel (b) of the in-sample 1-step ahead standardised forecast residuals obtained from Kalman filtering of the applied DBSTAR model C. For a well-specified model (7.2), the standardised forecast errors are serially independent and normally distributed. From Panel (a), the residuals seem to follow approximately a normal distribution, since the density curve looks symmetric around zero and with no heavy tails. It is worth pointing out that there are more positive residuals than negative ones. It is clearer if we look at the bar between 0 and 1 which is higher than between -1 and 0. The model seems to under forecast the electricity load fairly more often than over forecast, so the residuals present such a feature. Overall, the model presents well-behaved in-sample 1-step ahead standardised forecast residuals, which is satisfactory to show that the DBSTAR model C does capture all dynamics well with respect to each component individually.

Since the main focus of this application is on 1-step ahead forecasting, concentra-



(a)



(b)

Figure 7.13: In-sample Panel (a) histogram together with normal density function (blue curve) and Panel (b) boxplot of 1-step ahead standardised forecast residuals.

tion is on the dynamic features of the forecast residuals. A particularly interesting diagnostic tool is the set of boxplots of the in-sample 1-step ahead standardised forecast residuals per each hour of the day as illustrated in Panel (b). This figure complements the analysis from Panel (a), indicating the periods of the day when the model forecasts are (are not) accurate, in accordance with the variability of the electricity load as shown in Figures 7.2 and 7.8.

Generally, the variability of the standardised residuals are low at the first 4 hours of the day, presenting a considerable increase at 5AM, with emphasis at 6AM and remains high until 9AM, when it drops back to low levels. The variability increases again at around lunchtime and decreases back in the afternoon. Between 5PM and 8PM the residuals present very large variability. And after 9PM it drops back to low levels.

In summary, the peak times in the morning and in the evenings are more difficult to forecast. Overall, the dynamic structure of the standardised residuals of the DBSTAR model C is reasonably satisfactory.

7.6 Forecasting performance

In this section, the out-of-sample analysis is carried out in order to evaluate the forecasting performances of competing DBSTAR models. The observation period starts at the first hour on 1 June 2010 and finishes at the last hour on 30 June 2010, therefore 1 month of hourly data. This period presents not only standard weekdays and weekends but also particular non-standard behaviours such as holidays and some special events, such as the Brazilian Valentines' Day and the World Cup matches.

Table 7.3: Out-of-sample mean absolute percentage error (MAPE) and log-likelihood of predictive distribution (LPL) of each DBSTAR model and the random walk (RW) model

Model	MAPE(%)	LPL
A	1.99	-5983.59
B	2.19	-5990.85
C	**1.98**	**-5982.78**
RW	**4.09**	-

When evaluating model forecasting performance in this application, it is essential to take into account not only the performance of point forecasts but also the rigor of their accuracy. Therefore, the LPL defined in equation (3.4), is used to determine the accuracy of the joint forecast distribution for the out-of-sample period under investigation. Recall that the larger the value of the LPL, the more evidence in favour of the corresponding model. In addition, the most famous measure of accuracy of a time series model in this area is also analysed. It refers to mean absolute percentage error (MAPE), expressed as,

$$MAPE = 100 \sum_{t=1}^{T} \frac{|e_t/y_t|}{T}$$

Table 7.3 presents the out-of-sample measures, the mean absolute percentage error MAPE (in %) and the LPL, both used to assess forecasting performance. The three DBSTAR models show very similar values of MAPE and LPL with a slightly better forecast accuracy for Model C, as it has the smallest value of MAPE and the largest LPL, as identified in black bold face. During the in-sample period, the models' adaptive behaviours are very fast and precise that they present very small forecasting errors. Analogous performance is obtained during out-of-sample period.

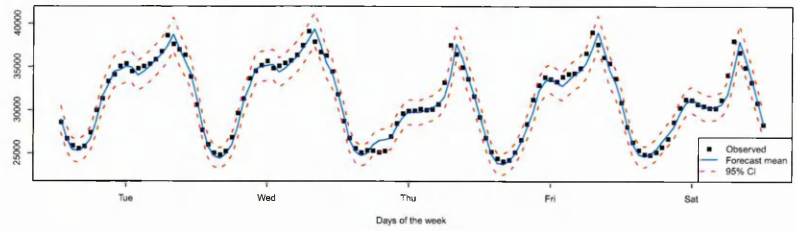
Table 7.3 also presents the out-of-sample MAPE (in %) of the random walk (RW) model used as a benchmark model. For a review, see Hamilton (1994). This non-stationary forecasting model is largely used in studies in many areas, and it is defined as

$$Y_t = y_{t-1} + \epsilon_t \tag{7.8}$$

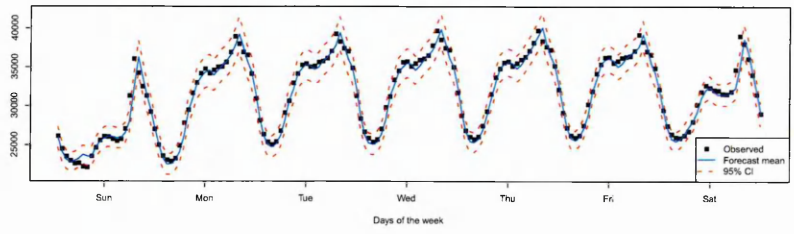
where ϵ_t is a white noise sequence with zero mean and variance σ^2. From this benchmark model, the 1-step ahead forecast of electricity load is simply the observed electricity load from the immediate previous hour. Its MAPE is roughly two times larger than the MAPE from any DBSTAR model. It indicates that every time a considerably large increase or decrease in the electricity load from an hour to the next occurs, this RW model will produce an inaccurate forecast.

Figure 7.14 illustrates both the observed series in the out-of-sample period, the 1-step ahead forecast means and the 95% credible interval of Model C; Panel (a) from hour 1 on 1 June 2010 to hour 24 on 5 June 2010, Panel (b) from hour 1 on 6 June 2010 to hour 24 on 12 June 2010, Panel (c) from hour 1 on 13 June 2010 to hour 24 on 19 June 2010, Panel (d) from hour 1 on 20 June 2010 to hour 24 on 26 June 2010 and Panel (e) from hour 1 on 27 June 2010 to hour 24 on 30 June 2010.

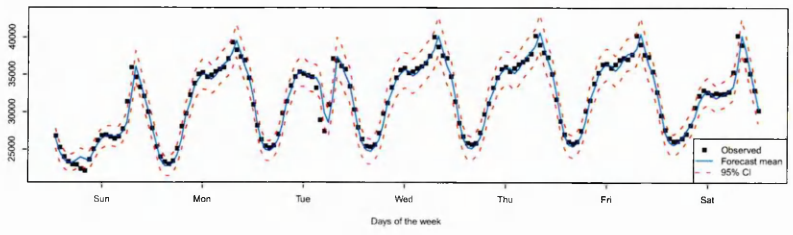
The past data up to the last hour of 31 May 2010 were used to make inference about the future - the whole month of June 2010. From the Kalman Filter algorithm, as described in Section 3.2.4, the forecast of the first observation in the out-of-sample period in Panel (a), i.e., 1am of 1 June 2010, is taken from the posterior probability distribution at the last hour of the day before, i.e., the last observation of 31 May 2010 in the in-sample period. Recall that forecasts are produced as probability



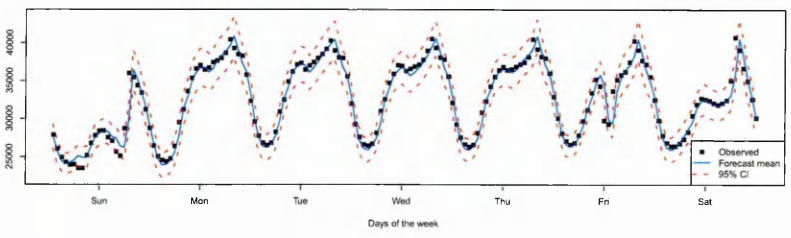
(a)



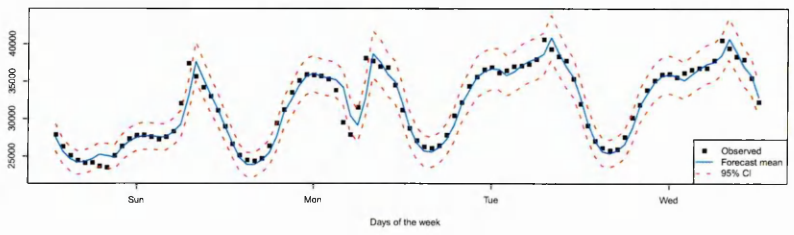
(b)



(c)



(d)



(e)

Figure 7.14: Out-of-sample 1-step ahead forecasting performances of Model C: Panel (a) from hour 1 on 1 June 2010 to hour 24 on 5 June 2010, Panel (b) from hour 1 on 6 June 2010 to hour 24 on 12 June 2010, Panel (c) from hour 1 on 13 June 2010 to hour 24 on 19 June 2010, Panel (d) from hour 1 on 20 June 2010 to hour 24 on 26 June 2010 and Panel (e) from hour 1 on 27 June 2010 to hour 24 on 30 June 2010.

distributions. After obtaining that forecast distribution, the first observation on 1 June 2010 becomes available to be inserted into the Kalman Filter algorithm and then the forecasting performance of the DBSTAR model can be assessed. Also, this observation updates the probabilities distributions from the previous hour and the forecast distribution for the next hour can be obtained. These steps are repeated at each time a new observation becomes available until the whole month of June 2010 is completely analysed and forecasted.

On Thursday 3 June 2010 was the *Corpus Christi* holiday. Note from Panel (a) that there is a change in the curve's shape, mainly with a drop in consumption most of the hours on that day. Similar changes were detected on the day after, Friday 4 June 2010, as it was a bridge-holiday, when compared with the other typical Fridays' electricity load in that month. The inclusion of a dummy component for holidays and bridge-holidays information in the modelling stage was rather important to capture the drop in electricity load and produce accurate forecasts. Without the inclusion of this component, the forecasting error on those special days were significantly higher since the electricity load decreases during holidays and bridge-holidays.

From Panel (b) the electricity load curve behaves well as there was no special days on that week, although on Friday 11 was the World Cup Opening Ceremony in South Africa. Perhaps because it was on TV at 9AM in Brazil on a typical working day, this event did not change the consumption's pattern. Moreover, on 12 June the Valentine's Day is celebrated in Brazil. In that year, it was on a Saturday and no particular changes in the pattern were observed either. Overall, the electricity load this week behaved as a standard weekdays' consumption from the in-sample period,

therefore the DBSTAR model could produce accurate forecast distribution for the entire week. This is the only week in the out-of-sample period when the electricity load behaves as standard since no atypical behaviour is identified and difficult to predict.

The third week of June 2010 is illustrated in Panel (c). On Tuesday 15, Brazil played North Korea for the World Cup [a] at 3:30PM. That special event drastically affected the electricity load pattern on that day, as watching football in Brazil is a big event, especially World Cup. Schools, universities, offices, and other institutions, generally opened part-time on that day, exempting people from study/work in the afternoon. That means that computers, lights, air-conditioners, fans, etc, were switched off, which reduced the consumption in the afternoon. In addition, it is a very common characteristic for Brazilians to get together to watch the matches. This could also explain the reduction of electricity load during the match time as people were together watching the match either at a bar, a restaurant, or a public square on a big screen, and therefore they were not spending electricity as in a typical day, such as ironing or having warm shower (as June is Winter time in Brazil), which requires more electricity than a TV does. Culturally, people stop doing their activities to fully concentrate on the match in front of the TV. Right after the match is finished, it is noticeable the rise in the load curve, when people get back to their activities.

In the following week, there were two Brazil matches. From Panel (d), analogous changes in the electricity load curve are identified on Sunday 20, as Brazil played Ivory Coast at 9AM, and on Friday 25, when Brazil played Portugal at 11AM.

In the last week of June 2010, Brazil played Chile on Monday 28 at 3:30PM and,

[a] http://www.fifa.com/worldcup/archive/southafrica2010/matches/groupstage.html

again, changes in the load pattern is identified as shown in Panel (e).

Accurate forecast distributions were obtained from the DBSTAR model, including the time when the electricity load drops during the World Cup matches, shown in Panels (c), (d) and (e). It can be explained by the fast degree of adaptability when a new observation becomes available to update all the probability distributions present in the model. In addiction, it does not take long to this new information be incorporated and forecast the load accurately. This is rather good results from the DBSTAR model since no extra components were included into the structure of the model to capture unexpected changes such as the World Cup matches.

Although June 2010 was a month full of special days for the electricity load, DBSTAR models presented a rather accurate forecasting performance as most observations lie within their respective forecasting credible intervals. During the in-sample period, the coefficients associated with the polynomial regression variables, the non-stationarity, the nonlinearity, all the cycles involved and the holidays are well specified and estimated. They all are adapted dynamically in time. The model therefore produces accurate forecast means at each hour during the out-of-sample period.

Accurate forecast variances were also produced by the DBSTAR model. The 95% credible intervals cover most of the observed series the entire period. It is worth noting that the credible intervals of the forecasts are wider at peak times (from 6PM to 9PM) than during off peak hours, regardless of the day. This analyse in out-of-sample period follows similar patterns identified during in-sample period, shown in Figure 7.8. Higher variability is observed at peak times and, consequently,

it is marginally more difficult for the models to predict the load in that period. It is a rather important result from DBSTAR models as the variance is allowed to change in time adapting the variability present in the electricity load, so the observed points at those times lie within the intervals.

7.7 Discussion

With rather satisfactory results shown in previous Chapter 6 obtained from proposed DBSTAR models, in this chapter, we focused on applying three DBSTAR models to a Brazilian electricity load high frequency data set. We aimed to capture the appropriate characteristics of the data in order to represent the underlying process with extensions of proposed Taylor and Splines DBSTAR models.

Temperature was used as the transition variable due to the nonlinear association between temperature and electricity load. The electricity load increases when temperature is increased but in a nonlinear shape, with an influence on load much larger with warmer temperatures. The shape of the electricity load curve at peak times varies according to the season. During winter, there is a sharp aspect in the curve in the evenings which is not present in the summer curve due to the BRST that encourages the use of natural lights early in the evenings, the peak times.

It was detected that electricity load is a long-term non-stationary process. Therefore, the models included a stochastic trend component for capturing explicitly the non-stationarity. This analysis ensures that both local and global non-stationarities are appropriately modelled by the DBSTAR approach with positive growth at non-constant rates.

A within-day cyclical pattern was detected from the similarity of the electricity load from one day to the next, particularly on weekdays. Similarly, a within-week cyclical pattern of the electricity load was also identified. Accordingly, two periodic components were used to represent those cycles.

A dummy covariate for holidays and bridge-holidays was incorporated into DBSTAR models structure as those special days affect the electricity load negatively.

Kalman filtering was used for obtaining the posterior distributions of the parameters as well as the forecast distributions of the electricity load. In-sample and out-of-sample analyses were carried out. During the in-sample period, the importance of those components in the models was reinforced. The time-dependent structure of DBSTAR models enables the parameters to adapt to periods with changes in patterns or special events, such as holidays. Overall, the models present well behaved in-sample 1-step ahead standardised forecast residuals, which gives strong evidence that DBSTAR models capture all dynamics present in the data well due to the use of explicit components individually. The result is a more parsimonious model, appropriate for time series with very short time intervals, since it avoids the inclusion of more parameters to capture these specific effects.

Out-of-sample analysis was carried out in order to evaluate the forecasting performances of competing DBSTAR models. The whole month of June 2010 was the period used to check the performances of the models. It contains special days with challenging patterns for the models to capture their electricity load's characteristics.

DBSTAR models presented rather accurate forecasting performance the entire month as most observations lie within their respective forecasting credible inter-

vals. On Thursday 3 June 2010 was the *Corpus Christi* holiday and due to the use of a dummy variable, DBSTAR models detected the change in the curve's shape, mainly with a drop in consumption for most of the hours on that day, and produced satisfactory forecasting results.

In that month, Brazil played five World Cup matches. Watching football in Brazil is a big event, especially World Cup, which affects drastically the electricity load pattern on those days. Culturally, people stop doing their activities to be fully concentrated on the matches in front of the TV. This could well explain the reduction of electricity load during the match times as consumers may not be using electricity as in a typical day.

The conclusion is that DBSTAR models adapt well to non-stationary nonlinear time series with particular patterns such as periodic repetitions, holidays and special events.

Chapter 8

Conclusion

8.1 Summary and main contributions

The main goal of this research is to propose Gaussian Dynamic Bayesian Smooth Transition Autoregressive (DBSTAR) models for non-stationary nonlinear autoregressive time series processes. DBSTAR models consider analytical approximations for STAR models based on Dynamic Linear Models of West and Harrison (1997) as alternative to both the classical STAR of Chan and Tong (1986a) and CBSTAR models of Lopes and Salazar (2005).

Classical STAR models use ordinary least squares for estimating the autoregressive coefficients and nonlinear least squares for the parameters within the transition function. A problem with this approach is that it requires initial values to start off the algorithm to be chosen and convergence can be very slow and not guaranteed. Lopes and Salazar (2005) developed CBSTAR models which use MCMC algorithms due to the loss of analytical tractability in calculating posterior distribu-

tions of parameters. Therefore, both the classical STAR and CBSTAR models are not generally appropriate for real-time applications, especially high frequency data.

Taylor series approximations and splines functions, such as Truncated Power Basis and B-splines, were used as solutions for the transition function $\pi(s_t; \underline{\psi}_t)$, such as the logistic or exponential functions. DBSTAR models sequentially update their dynamic parameters in time via Kalman filtering. This solution estimates the parameters in analytical closed form sequentially in time. In this way, we can avoid some computational problems associated with classical STAR and CBSTAR models, such as different convergence for different starting values.

At each time t, the first two moments of the original parameters can be obtained from the first two moments of the posterior Student-t distribution of the state vector by solving one system of polynomial equations. It is worth mentioning that this analysis cannot be done using either classical STAR or CBSTAR models as they are static models and none of their constant parameters has evolution in time.

Due to the Bayesian formulation, proposed DBSTAR models also has the advantage over classical STAR models of allowing formal inputs and interventions from experts, where appropriate. In addition, DBSTAR models can be applied to time series data sets when non-stationarity is present. It is not necessary to difference the time series to achieve stationarity.

DBSTAR models can also be applied to data in the presence of heteroscedasticity based on a conjugate analysis. The assumption of constant observational variance might be unrealistic depending on the application. Unlike the classical STAR and CBSTAR model formulations, DBSTAR models incorporate a variance discount

technique. We assume that the observational variance may change but only slowly and steadily over time.

Notice that classical STAR models are a particular case of DBSTAR models for a constant observational variance $V_t = \sigma^2$, fixed autoregressive coefficients $\underline{\theta}_t = \underline{\theta}$, an identity matrix set for the evolution of the states $\mathbf{G}_t = \mathbf{I}$, a null state covariance matrix $\mathbf{W}_t = \mathbf{0}$ (equivalently, for $\delta_W = 1$) and fixed smoothing parameter, $\gamma_t = \gamma$ and, consequently, $\pi(s_t; \gamma_t, c) = \pi(s_t; \gamma, c)$. Recall that classical STAR models estimate the parameters using the whole data set, so for the purposes of comparison, the modeller should run both the Kalman filter and the Kalman smoother algorithms to compare fits.

Many stationary models from both classical and Bayesian inferences (with the use of MCMC algorithms) eliminate the trend by differencing the series $(Y_t - Y_{t-1})$ as the dynamic changes cannot be handled by them. For instance, Lopes and Salazar (2005) differenced the US IPI (Industrial Production Index in the US) data four times to achieve stationarity and remove possible seasonality before modelling it using the CBSTAR model. However, DBSTAR models are suitable for modelling global and local non-stationary processes with additional model components to address those features. Furthermore, either linear trend or nonlinear time trend curves can be accommodated into their formulations.

DBSTAR models can be extended for modelling observed cyclical behaviour in terms of cyclical components added to the model structure. A periodic DBSTAR model is defined as a DBSTAR with an explicit component for a cycle with h harmonics. Fourier form representations of a cycle are considered as sine/cosine waves

which provide economic characterisation on parameters. In general, we use lower autoregressive order (less autoregressive coefficients) in the model when more weights are given to the past values of the dependent variable. The guaranteed parsimony of periodic DBSTAR models is balanced by larger amplitudes in the autoregressive coefficients. This is an advantage of periodic DBSTAR models over the standard DBSTAR, classical STAR and CBSTAR approaches, for modelling time series in the presence of cyclical behaviour.

DBSTAR models can be extended straightforwardly to investigate the dependence of a variable Y_t, not only on the past values of the series, but also on the values of other time series predictor variables $\underline{x}_t$. Therefore, autoregression and regression are both investigated in this formulation. The vector $\underline{x}_t$ may accommodate (i) exogenous time series observed at same time points as the dependent variable Y_t, (ii) lagged exogenous time series, i.e., past values of other time series variables ($\underline{x}_{t-d}$, where d is the delay parameter) and/or (iii) dummy variables. Any number of predictor variables can be added to the model structure.

Formulations of DBSTAR models were applied to the well-known Canadian lynx data in order to validate the proposals by comparing their fitting performances with the performances of both the classical STAR and CBSTAR models. The analyses during in-sample period suggest that the time-dependent parameters DBSTAR models presented fast and precise adaptive behaviour producing very small fitting errors compared to static DBSTAR models. The cycle components changing throughout the observed time series period were appropriately modelled by them. We concluded also that standard models try to capture the cycle using high autoregressive orders

while periodic models use low autoregressive orders plus periodic components to model the cycle explicitly. Nonetheless, in the out-of-sample period, the forecast means are, overall, accurate but the forecast uncertainties are overestimated for time-dependent DBSTAR models.

Compared to the classical STAR and the CBSTAR approaches, DBSTAR models showed improved fitting performances. On the other hand, they present large number of parameters to be sequentially updated over time. However, the algorithms run in just a few seconds for each model configurations. Hence, proposed DBSTAR models are promising for real-time applications.

With acceptable results shown in the application to the Canadian lynx data set, extended versions of DBSTAR models were applied to a Brazilian electricity load high frequency data set. With the inclusion of components in the models structures, we aimed to capture the appropriate characteristics from the data in order to represent the underlying process.

DBSTAR models performed reasonably well during both in-sample and out-of-sample periods. During the in-sample period, the importance of those components in the models was reinforced. Overall, the models presented well behaved in-sample 1-step ahead standardised residuals in both periods, which gives strong evidence that DBSTAR models capture all dynamics well present in the data due to the use of explicit components individually. During the out-of-sample period, DBSTAR models presented a fairly accurate forecasting performance as most observations lie within their respective forecasting credible intervals.

The conclusion is that DBSTAR models adapt well to non-stationary nonlinear

time series with particular patterns such as periodic repetitions, holidays and special events. The result is a more parsimonious model, very appropriate for very short time intervals data set, since it avoids the inclusion of more parameters to capture these specific effects.

8.2 Future research

There are some improvements that the current versions of DBSTAR models can address in order to make this class of models more suitable for non-stationary nonlinear time series processes.

In Section 5.4, we proposed DBSTAR models to accommodate more than 2 regimes, called multiple regimes DBSTAR models. Those models were not applied to either data set in chapters 6 and 7, as simpler versions of DBSTAR seemed to work fairly acceptably. We suggest further investigation that may point out that those data sets require to be modelled by multiple regime approaches. For instance, the electricity load application that used temperature as a transition variable could be the case when DBSTAR models are split into more than two regimes, such as a regime for temperature below 15^oC, another for temperature between 15^oC and 25^oC, and a third model when temperature is 25^oC and 35^oC and finally for very hot days with temperature over 35^oC.

In case they do not require more than 2 regimes in their models structure, multiple regime models should then be applied to other data sets for further evaluation. One idea is to apply the multiple regimes DBSTAR models to the same data set used in van Dijk and Franses (1999) and compare their fitting performances with

classical multiple regimes STAR models.

Another extension to be proposed in the future is the multivariate DBSTAR models for nonlinear time series with more than one dependent variable. This approach will estimate all parameters simultaneously. A point to note is that going from the univariate to the multivariate version will require the estimation of more parameters. It should therefore be aimed to find parsimonious formulations of the multivariate model as the univariate versions already present large number of parameters.

One example of an application of this suggested multivariate DBSTAR is in the electricity load data set. The load $(\mathbf{Y}_t)$ would be a 24-dimensional vector instead of a scalar at each time t to model concomitantly the 24 loads of each day. It means that the relationships between loads at different hours would be affected by the present and past of the loads, in addition to the transition variable temperature. It would be easier to work out k-step ahead forecasts for the next 24 hours. However, multivariate models are far more complicated, since they have big problems of how to estimate the covariance structure, see Barbosa and Harrison (1992) and West and Harrison (1997).

As an aside, multivariate DBSTAR models could be proposed to investigate the presence of co-integration in nonlinear processes. For details on co-integration see (Granger, 1988) and (Engle and Yoo, 1987). Co-integration is an important concept, mainly investigated in empirical macroeconomics, and relates to the number of stochastic trends which are present in a set of time series. Therefore, co-integration would be modelled by trend components incorporated into the structure of multi-

variate DBSTAR models.

It is also recommended to extend the current versions to non-normal DBSTAR models. There are countless time series that are modelled by Gaussian approaches whilst they should actually be modelled by non-Gaussian distributions. One example is the Canadian lynx application that is count data and should be modelled by a poisson process. Hence, a poisson DBSTAR model should be proposed to that case. This extended non-normal DBSTAR proposals seem to be straightforwardly implemented if the distribution belongs to the exponential family, since there are algorithms in the literature to estimate the parameters for this case (see West *et al.* (1985)).

Finally, a fully Bayesian DBSTAR approach that can be used in real-time application would use Sequential Monte Carlo methods, such as, Particle filtering, for updating all the parameters in the model. It is worth mentioning that the Kalman filter algorithm is a particular case of Particle filtering. For a review on Particle filtering methods, see Lopes and Tsay (2011) and Young (2011).

There are actually two possible ways of proposing DBSTAR models using Particle filtering methods rather than the standard Kalman filter: (i) the structure of DBSTAR models is kept as proposed in this thesis, i.e., take the polynomial forms using approximation methods (Taylor series or splines) written in the DLM form, and just replace the estimation methods, i.e., use one of the Particle Filtering algorithms to assess the posterior distribution of the parameters rather than the standard Kalman filter algorithm used in this thesis; or (ii) take the STAR model with parameters evolving in time from equation (4.2) and directly rewrite it into the DLM form

without using any approximation methods for the transition function. That means rewriting a nonlinear model into a linear approach. The standard Kalman filter algorithm can no longer be used for assessing the posterior and forecast distributions in this case. However, Particle filtering methods should be implemented instead. In both methods, there will be no need for using model selection criteria for choosing values of parameters which cannot be accommodated in the state vector, such as the autoregressive order p, the Taylor series degree r or the polynomial degree n of splines, and others, since Particle filtering methods fully account for all model parameter uncertainties.

Knowing that there is no impeccable model, we pursue for further research in non-stationary nonlinear time series in the future.

Bibliography

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19:716–723.

Atkinson, K. E. (1988). *An introduction to Numerical Analysis.* John Wiley & Sons.

Bacon, D. W. and Watts, D. G. (1971). Estimating the transition between two intersecting straight lines. *Biometrika*, 58:525–534.

Barbosa, E. and Harrison, J. (1992). Variance estimation for multivariate dynamic linear models. *Journal of Forecasting*, 11:621–628.

Bates, D. M. and Watts, D. G. (1980). Relative curvature measures of nonlinearity. *Journal of the Royal Statistical Society. Series B (Methodological)*, 42:1–25.

Bauwens, L., Lubrano, M., and Richard, J. (1999). *Bayesian Inference in Dynamic Econometric Models.* Oxford University Press.

Box, G. and Jenkins, G. (1970). *Time Series Analysis.* Holden–Day.

Box, G. E. P. and Tiao, G. C. (1965). A change in level of a non-stationary time series. *Biometrika*, 58:181–192.

Bretthorst, G. L. (1988). *Bayesian Spectrum Analysis and Parameter Estimation.* Springer-Verlag.

Broemeling, L. D. (1985). *Bayesian Analysis of linear models.* Decker.

Broemeling, L. D. and Cook, P. (1992). Bayesian analysis of threshold autoregressions. *Communications in Statistics - Theory and Methods*, 21:2459–2482.

Campbell, E. P. (2004). Bayesian selection of threshold autoregressive models. *Journal of Time Series Analysis*, 25:467–482.

Chan, K. and Tong, H. (1986a). On estimating thresholds in autoregressive models. *Journal of Time Series Analysis*, 7:179–190.

Chan, W. S. and Tong, H. (1986b). On tests for non-linearity in time series analysis. *Journal of Forecasting*, 5:217–228.

Chen, C. and Lee, J. (1995). Bayesian inference of threshold autoregressive models. *Journal of Time Series Analysis*, 16:483–492.

Chen, C. W. (1998). A bayesian analysis of generalized threshold autoregressive models. *Statistics & Probability Letters*, 40:15–22.

Chernoff, H. and Zacks, S. (1964). Estimating the current mean of a normal distribution which is subjected to changes in time. *The Annals of Mathematical Statistics*, 35:999–1018.

de Boor, C. (1978). *A Practical Guide to Splines.* Springer.

Dordonnat, V., Koopman, S., Ooms, M., Dessertaine, A., and Collet, J. (2008). An hourly periodic state space model for modelling french national electricity load. *International Journal of Forecasting*, 24:566–587.

Dufrénot, G., Mignon, V., and Péguin-Feissolle, A. (2003). Business cycles asymmetry and monetary policy: a further investigation using mrstar models. *Economic Modelling*, 21:37–71.

Elton, C. and Nicholson, M. (1942). The ten-year cycle in numbers of the lynx in canada. *Journal of Animal Ecology*, 11:215–244.

Engle, R. and Yoo, B. (1987). Forecasting and testing in co-integrated systems. *Journal of Econometrics*, 35:143–159.

Ferreira, P. E. (1975). A bayesian analysis of a switching regression model: Known number of regimes. *Journal of the American Statistical Association*, 70:370–374.

Gelfand, A. E. and Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85:398–409.

Gelman, A., Carlin, B., and Rubin, H. S. D. (1995). *Bayesian Data Analysis.* Chapman & Hall.

Geweke, J. and Terui, N. (1993). Bayesian threshold autoregressive models for nonlinear time series. *Journal of Time Series Analysis*, 14:441–454.

Gilks, W. R., Richardson, S., and Spiegelhalter, D. (1996). *Markov Chain Monte Carlo in Practice.* Chapman & Hall.

Good, I. J. (1985). Weight of evidence: A brief survey. In Bernardo, J., DeGroot, M., Lindley, D., and Smith, A., editors, *Bayesian Statistics 2*, pages 249–270. Elsevier Science Publisher B V, North–Holland, Amsterdam, and Valencia University Press.

Granger, C. W. J. (1988). Some recent development in a concept of causality. *Journal of Econometrics*, 39:199–211.

Green, P. J. (1995). Reversible jump markov chain monte carlo computation and bayesian model determination. *Biometrika*, 82:711–732.

Hamilton, D. C., Watts, D. G., and Bates, D. M. (1982). Accounting for intrinsic nonlinearity in nonlinear regression parameter inference regions. *The Annals of Statistics*, 10:386–393.

Hamilton, J. (1994). *Time Series Analysis.* Princeton University Press.

Hinkley, D. V. (1969). Inference about the intersection in two-phase regression. *Biometrika*, 56:495–504.

Hudson, D. J. (1966). Fitting segmented curves whose join points have to be estimated. *Journal of the American Statistical Association*, 61:1097–1129.

Jeffreys, H. (1961). *Theory of Probability.* Oxford University Press.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Journal of Basic Engineering (Transactions ASME)*, 82:35–45.

Kander, Z. and Zacks, S. (1966). Test procedures for possible changes in parameters

of statistical distributions occurring at unknown time points. *The Annals of Mathematical Statistics*, 37:1196–1210.

Kim, M. S. (2013). Modeling special-day effects for forecasting intraday electricity demand. *European Journal of Operational Research*, 230:170–180.

Koop, G. and Potter, S. M. (1999). Bayes factors and nonlinearity: Evidence from economic time series. *Journal of Econometrics*, 88:–.

Koop, G. and Potter, S. M. (2003). Bayesian analysis of endogenous delay threshold models. *Journal of Business & Economic Statistics*, 21:93–103.

Kutner, M., Nachtsheim, C. J., Neter, J., and Li, W. (2005). *Applied linear statistical models.* Mc Graw Hill.

Lawrance, A. J. and Kottegoda, N. T. (1977). Stochastic modelling of river flow time series. *Journal of the Royal Statistical Society. Series A (General)*, 140:1–47.

Lopes, H. F. and Salazar, E. (2005). Bayesian model uncertainty in smooth transition autorregressions. *Journal of Time Series Analysis*, 27:99–117.

Lopes, H. F. and Tsay, R. S. (2011). Particle filters and bayesian inference in financial econometrics. *Journal of Forecasting*, 30:168–209.

Lundbergh, S., Teräsvirta, T., and van Dijk, D. (2003). Time-varying smooth transition autoregressive models. *Journal of Business & Economic Statistics,*, 21:104–121.

Luukkonen, R., Saikkonen, P., and TerasvirtaTeräsvirta, T. (1988). Testing linearity against smooth transition autoregressive models. *Biometrika*, 75:491–499.

Mendonça, A. F. and Dahl, C. (1999). The brazilian electrical system reform. *Energy Policy*, 27:73–83.

Moita, R. M. (2008). Entry and externality: Hydroelectric generators in brazil. *International Journal of Industrial Organization*, 26:1437–1447.

Moran, P. A. P. (1953). The statistical analysis of the canadian lynx cycle, i: Structure and prediction. *Australian Journal of Zoology*, 1:163–173.

Mustafi, C. K. (1968). Inference problems about parameters which are subjected to changes over time. *The Annals of Mathematical Statistics*, 39:840–854.

Ortega, J. M. and Rheinboldt, W. C. (1970). *Iterative solution of nonlinear equations in several variables.* Duxbury.

Page, E. S. (1954). Continuous inspection schemes. *Biometrika*, 41:100–115.

Page, E. S. (1955). A test for a change in a parameter occurring at an unknown point. *Biometrika*, 42:523–527.

Page, E. S. (1957). On problems in which a change in a parameter occurs at an unknown point. *Biometrika*, 44:248–252.

Pardo, A., Meneu, V., and Valor, E. (2002). Temperature and seasonality influences on spanish electricity load. *Energy Economics*, 24:55–70.

Perobelli, F. S. and de Oliveira, C. C. C. (2013). Energy development potential: An analysis of brazil. *Energy Policy*, 59:683–701.

Pole, A., West, M., and Harrison, J. (1994). *Applied Bayesian Forecasting and Time Series Analysis.* Chapman & Hall.

Priestley, M. (1988). *Non-linear and Non-stationary Time Series Analysis.* Academic Press.

Quandt, R. E. (1958). The estimation of the parameters of a linear regression system obeying two separate regimes. *Journal of the American Statistical Association*, 53:873–80.

Quandt, R. E. (1960). Tests of the hypothesis that a linear regression system obeys two separate regimes. *Journal of the American Statistical Association*, 55:324–330.

Ramsay, J. O., Hooker, G., and Graves, S. (2009). *Functional Data Analysis with R and MATLAB.* Springer.

Robison, D. E. (1964). Estimates for the points of intersection of two polynomial regressions. *Journal of the American Statistical Association*, 59:214–224.

Rosa, L. P. and Lomardo, L. L. B. (2004). The brazilian energy crisis and a study to support building efficiency legislation. *Energy and Buildings*, 36:89–95.

Ruppert, D., Wand, M. P., and Carroll, R. J. (2003). *Semiparametric Regression.* Cambridge University Press.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6:461–464.

So, M. K. P. and Chen, C. W. S. (2003). Subset threshold autoregression. *Journal of Forecasting*, 22:49–66.

Soares, L. and Souza, L. (2006). Forecasting electricity demand using generalized long memory. *International Journal of Forecasting*, 22:17–28.

Spiegelhalter, D., Best, N., Carlin, B., and van der Linde, A. (2002). Bayesian measures of model complexity and fit (with discussion). *Journal of The Royal Statistical Society*, 64:583–639.

Stenseth, N., Falck, W., Chan, K.-S., Bjrnstad, O., O'Donoghue, M., Tong, H., Boonstra, R., Boutin, S., Krebs, C., and Yoccoz, N. (1998). From patterns to processes: Phase and density dependencies in the canadian lynx cycle. *Proceedings of the National Academy of Sciences of the United States of America*, 95:15430–15435.

Teräsvirta, T. (1994). Specification, estimation, and evaluation of smooth transition autoregressive models. *Journal of the American Statistical Association*, 89:208–218.

Teräsvirta, T. (2005). Forecasting economic variables with nonlinear models. *Working Paper Series in Economics and Finance*, 598:1–55.

Tong, H. (1977). Contribution to the discussion of the paper entitled *Stochastic modelling of river flow time series* by A. J. Lawrance and N. T. Kottegoda. *Journal of the Royal Statistical Society. Series A (General)*, 140:34–35.

Tong, H. (1978). *On a threshold model.* In Patern Recognition and Signal Processing.

Tong, H. (1983). *Threshold models in non-linear time series analysis.* Springer.

Tong, H. (1990). *Non-linear Time Series: A Dynamical System Approach.* Oxford University Press.

Tong, H. and Lim, K. S. (1980). Threshold autoregression, limit cycles and cyclical data. *Journal of the Royal Statistical Society. Series B (Methodological)*, 42:245–292.

Tsay, R. S. (1986). Nonlinearity tests for time series. *Biometrika*, 73:461–466.

Tsay, R. S. (1989). Testing and modeling threshold autoregressive processes. *Journal of the American Statistical Association*, 84:231–240.

Tsay, R. S. (1991). Detecting and modeling nonlinearity in univariate time series analysis. *Statistica Sinica*, 1:431–451.

van Dijk, D. and Franses, P. H. (1999). Modeling multiple regimes in the business cycle. *Macroeconomic Dynamics,*, 3:311–340.

van Dijk, D., Strikholm, B., and Teräsvirta, T. (2003). The effects of institutional and technological change and business cycle fluctuations on seasonal patterns in quarterly industrial production series. *Econometrics Journal*, 6:79–98.

van Dijk, D., Teräsvirta, T., and Franses, P. H. (2002). Smooth transition autoregressive models–a survey of recent developments. *Econometric Reviews*, 21:1–47.

Veselovsky, I. and Tarsina, M. (2002). Intrinsic nonlinearity of the solar cycles. *Advances in Space Research*, 29:417–420.

West, M. (1986). Bayesian model monitoring. *Journal of Royal Statistical Society B*, 48:70–78.

West, M. and Harrison, J. (1997). *Bayesian Forecasting and Dynamic Models.* Springer.

West, M., Harrison, J., and Migon, H. (1985). Dynamic generalized linear models and bayesian forecasting. *Journal of the American Statistical Association*, 80:73–83.

Whittle, P. (1954). The statistical analysis of a seiche record. *Journal of Marine Research*, 13:76–100.

Young, P. C. (2011). Gauss, kalman and advances in recursive parameter estimation. *Journal of Forecasting*, 30:104–146.

Zhang, W. G. and Duda, T. F. (2013). Intrinsic nonlinearity and spectral structure of internal tides at an idealized mid-atlantic bight shelf break. *Journal of Physical Oceanography*, 43:2641–2660.